

New ways with research support

A British committee urges more openness on and competition for research funds. Its methods deserve a serious trial.

HERE is a question from the final examination paper of the course in Public Policy Priorities (PPP) recently introduced at the University of the Sticks:

QUESTION: Devise a scheme for spending \$1,000 million of public money each year on scientific research, paying particular attention to (a) your government's need to assert at all times that all initiatives have been or are about to be taken; (b) government's insistence that public money is always spent by those who can spend it best, but that no possible recipient has been overlooked; and (c) the need not to make routine demands on the time or judgement of civil servants. Draft terms of reference for all standing committees your scheme requires. (Time allowed: 30 minutes.)

The time allowance is not as niggardly as it seems. Students are not, after all, being asked to compare and contrast the roles of the French Centre National de la Recherche Scientifique and the US National Science Foundation in their respective environments, or to explain why the elected Science Council of Japan has run into trouble. A reasonably well-informed student has merely to write down on paper what he knows about the way his own government administers public support for non-military research. The time consuming part of the assignment is the drafting exercise and not because of any inherent difficulty but because (a), (b) and (c) require that there will be a great many committees. Merely listing them will take half the time allowed.

There is, however, just a chance that a year from now, British undergraduates will find the assignment more taxing. Since the Second World War, there have been inquiries into the organization of British civil science at intervals of roughly a decade, each of them accompanied by a restatement of the need that research should contribute more directly to prosperity and by the renaming of a few previously well-known institutions. By common consent, the formal reorganization of civil science has become unfashionable. Those concerned are exhausted by the upheavals of the past and they know that formal reorganization gets nowhere. Dissenters from that proposition should ask whatever happened to the great banner of "Research Applied to National Needs" reflected on the US National Science Foundation in the 1960s, or wish on how the British Science and Engineering Research Council, carved out of the pre-existing Department of Scientific and Industrial Research to be the chief source of support for university research, has ever since been trying to demonstrate its concern for the welfare of British industry. The new development, in Britain, is that people have begun to regard existing institutions as capable of change and that the people themselves have also changed, usually by retirement and succession.

Reform

The Morris report (page 383) on the relationship between British universities and research councils is an omen — or it may be. The committee responsible for this slim document was asked to tackle a long-standing source of discontent and friction — the proper relationship between four titularly autonomous research councils, jointly spending more than £500 million a year on basic and applied research — and the universities, which nominally receive more than £300 million a year to support research, but which are spending increasing proportions of that sum on the relief of the discomforts caused by tighter budgets. The committee has many

sensible things to say about how both kinds of institutions should reform themselves. The Advisory Board for the Research Councils, which instigated the inquiry, must now ensure that the councils live up to what is asked of them. But who will keep the universities up to the mark?

Morris offers a solution which is, in the context, daring. If what the British call the dual-support system is breaking down because many universities are spending public funds intended to support research for other purposes, why not arrange that they do not break the rules? Easier said than done, PPP students will say. For under the British system, universities are given an allocation of taxpayers' money each year but are never told how much is for teaching and how much for research. But if universities are not told explicitly what is expected of them, how can anybody complain if they choose to do something quite different with the money they are given? So why not make public how much of the annual budget is meant for research support?

This, the Morris solution, is both logical and simple but probably, on that account, unacceptable to the universities. For if the University Grants Committee (on whose recommendation these funds are distributed) were to make its purposes explicit, many universities would complain that their autonomy had been infringed, and there would be howls of protest that the allocation of research money was not equitable: some universities would be found to be dealt with more generously than others, thus openly denying the polite fiction that all universities are equal. The first principle of the public administration of science, that central government should not be bothered by day-to-day matters, would instantly be breached by a string of indignant questions in the House of Commons from people wanting to know why their local university had been short-changed.

Courage

That, fortunately, does not mean that the system should not be tried, but merely that it will take a courageous University Grants Committee to face up to its implications. And there is a problem. If an individual university research allocation is known explicitly, the arrangement should be accompanied by some means of making sure that what is settled now (in 1984-85?) is not immutable for the rest of time, so that universities will enjoy a sense that they can work their way up the league tables by their own efforts, perhaps by their skill in putting together research programmes that win support from other than public sources. A further advantage is that such a system could be the foundation for a bridge between universities and the polytechnics which now shadow them, for it is not impossible that similar rules could be devised for each of the two halves of public sector higher education. But the overriding advantage is that no new committees are needed — merely a change of practice by bodies that exist already.

That is the most practical of the Morris committee's radical proposals. The even more subversive notion that research council establishments carrying out work that might be done in universities should sometimes have to compete for support with academic scientists is unlikely to get far, at least in the short term. The machinery for such competition exists and consists of the research councils themselves. The difficulty is that the councils could force their own people to compete for research funds only with the greatest administrative difficulty, and in the knowledge



that failure would encumber them with the costs of under-employed people. Moreover the research councils, which often shouldered responsibility for research that universities were in danger of neglecting, would rightly protest that enforced competition for funds would often be unfair. Morris's objective in this connection seems to have been to find some way of applying common criteria to universities and to some of the large institutes supported by the research councils — the Natural Environment Research Council's Institute of Geological Sciences (budget: £28 million a year) and the Medical Research Council's National Institute for Medical Research are two obvious examples. PPP students will recognize that this problem arises wherever mission-oriented laboratories have embarrassingly outlived their terms of reference (cf. "national laboratories" in the United States). Perhaps a simpler solution is that large pieces of these establishments should be hived off to universities, with an appropriate transfer of funds when necessary. There again, no new machinery would be necessary.

But would British universities agree, especially when they are still preoccupied with the problems of protecting those among their own teaching staffs who have not leapt at the opportunity of early retirement on the generous terms agreed with the University Grants Committee? Not without consequential changes in their own organization which nevertheless would be valuable in themselves. One of the evil consequences of the past decade's short commons is that most British universities have become inward-looking places, more concerned with the defence of teaching posts and the stability of the student-staff ratio than with their wider social roles in education and with scholarship, some of which leads to industrial innovation. So academic snobbery, sometimes a thin disguise for self-defensiveness, has flourished. There would be much suspicion at the prospect of taking into universities groups of researchers from different environments even if the research councils and the grants committee were to ensure that universities would not suffer in the process. The need now, as the Morris committee appears to have spotted, is for the most sensible distribution of research people between universities and other establishments, and for more openly honest competition for what funds there are to spare. Nobody should complain if a few sacred cows have to be slaughtered in the process.

Organization

The traditional concept of how British universities should be run is one of these. People elsewhere will marvel that British universities enjoy a ratio of students to tenured members of staff which amounts to something like 10 to 1. Some institutions are better placed, perhaps only eight students to every member of staff. How is it that distinguished universities elsewhere, in the United States for example, manage with ratios which are often twice as great? The answer is, of course, that they are differently organized, with many more impermanent people, graduate students and more senior researchers, helping out with teaching. And the results are not always the disasters that British academics predict. One of these days, some British universities may recognize this as a model by means of which they can sustain a substantial research programme, teach more students and yet live within their budgets.

Most probably, however, that time is a long way off. The Morris committee complains in its report that British universities have been slow to follow the recommendation that they should create some mechanism by which their research could be evaluated internally and the research funds distributed by the grants committee shared out equitably. The delay is a measure of people's diffidence at knowing that their research may be discussed collectively — and their legitimate fear that internal mechanisms for distributing resources work politely only when there is an uncritical formula. Fair shares means equal shares is how the argument goes. That is a recipe but not in itself an excuse for doing nothing.

These are administrative matters. The most striking part of the Morris committee's report is not, however, administrative but hortatory. The document refers to the research community in affectionate terms, and by implication makes the obvious but for-

gotten point that the productivity of a group of people, motor-salesmen and academic researchers alike, depends to a large extent on their morale. Part of the trouble in Britain in the past ten years is that the research community has for too long been depressed by the apparently endless financial problems which afflict it, by the need to spend time on committees appointed to share out diminishing resources and by the general sense that even the brightest ideas will almost always come to nothing. Exhortation by itself will not cure this state of affairs, but encouragement and good example may. Morris offers the research councils the challenge of taking that responsibility on themselves. It is an interesting challenge, which may or may not be taken up. But the challenge carries a prize for anybody who can grasp it firmly. □

End the Sizewell agony

The British no longer build nuclear reactors but instead hold public inquiries about them.

SOME months ago, *Nature* offended some of its readers by saying that the British Government should abandon the public inquiry into the plan of the Central Electricity Generating Board to build just one pressurized water reactor at Sizewell in the county of Suffolk (301, 100; 1983). The most cogent of the reasons for this suggestion is largely constitutional. Under British planning legislation, people affected by major developments are rightly given a chance to protest, but the use of the planning procedures to decide whether there should be nuclear power stations at all, what prices the nationalized utilities should prudently pay for them and whether reactors of particular types are in some acceptable sense "safe" is a usurpation of the constitutional roles of the British Government, the House of Commons and public corporations such as the generating board. Now there is another case against this inquiry.

The proceedings thus fall into the category of observations that interfere with the system observed. While much of the argument about the building of one pressurized water reactor has turned on assertions that it could not repay the cost, it seems to have been forgotten that the inquiry itself will help to make that prophecy come true. Delay is notoriously a cause of cost escalation in the building of nuclear power plants, not merely in Britain but wherever they have been built. Taxpayers in Britain may bemoan the fees of the lawyers who argue the case about the Suffolk reactor, but the more serious concern should be the increased construction cost that will eventually be borne by themselves in another guise, as electricity consumers. For the cost of the design teams kept kicking their heels for two needless years, to say nothing of the cost of the interest on the work already carried out, will eventually appear on people's electricity bills, whatever the inquiry decides. So should the proceedings be allowed to continue?

Paradoxically, they are taking place at all only because the Secretary of State for Energy in the last British government but one, Mr Tony Benn, promised that they would. The first Thatcher government (1979–83) began by announcing a brave programme of nuclear construction and apparently thought it prudent to let the inquiry take its generic course. But now, it seems, the electricity utility has given up the hope that the first pressurized water reactor would begin a series, with later versions benefiting from the savings that would come from serial design. For the time being, there is just not enough room in the electricity generating network for such luxuries. In the circumstances, it would be entirely sensible that the terms of reference of the inquiry should be restricted to the local planning issues. The present government, which has most to lose from the derogation of its power to conduct energy policy as it thinks fit, and which has not been slow to override more formal judicial processes in other fields (as with its private deal last week to exempt the London Stock Exchange from its own legislation on the restraint of trade) should find some way, during the summer, of ending the agony over the Suffolk reactor. □

UK research support

Committee advocates more competition

A RADICAL change in the mechanism for supporting British universities is suggested in a report* of a committee on academic research, published last week. The proposal is that the University Grants Committee should separately identify in its annual grants to British universities the sums of money made available for research.

The objective is to discourage universities from spending research funds for other purposes, but it runs directly counter to the traditional mystique of the University Grants Committee, which disburses the funds and which has always resisted suggestions that its decisions derive from an explicit formula.

The committee, which also has pointed comments on the ways in which British research councils conduct their affairs, was set up last year by the Advisory Board for the Research Councils to look into the relationship between research councils and universities.

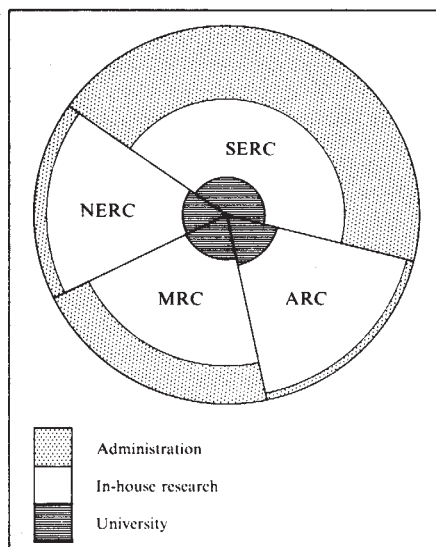
The chairman, Mr J.R.S. Morris (also chairman of Brown and Root (UK) Ltd) is a newcomer to these circles but has a reputation for being able to make up his mind. It may be significant that the signatories of the report include Sir David Phillips, the new chairman of the advisory board, and Sir Peter Swinnerton-Dyer, the new chairman of the University Grants Committee. Neither appears to have dissented from the recommendations.

That may be because the committee has ducked the most contentious issue on its agenda — that of how research councils should divide their resources between in-house research and support for research in universities. The report does, however, quote complaints from several universities that research councils are over-protective of their in-house work.

The proposal that the research element in public support for individual universities should be separately known is intended to ensure that universities do not skimp on research support when budgets are tight. Evidence that this temptation is real is provided by some of the representations to the committee by universities, and is the reason why some research councils have recently been compelled to augment research grants to academics by payments to meet the cost of equipment that would normally have been provided by universities.

But the proposal flatly contradicts one of the conclusions of a working party on the British dual-support system under Sir

Alec Merrison. It is likely to cause most offence among universities, especially those guilty of skimping on research, on the grounds that their autonomy would thus be further undermined. The Morris committee also recommends that the research component of university recurrent grants should be allocated to different departments by a research committee — and then points out that universities have been slow to set up machinery along these lines although urged to do so a year ago.



The support given by research councils for in-house and university research. ARC, Agricultural Research Council; MRC, Medical Research Council; NERC, Natural Environment Research Council; SERC, Science and Engineering Research Council. Most in-house spending at SERC is at service laboratories and on international subscriptions. Data from the Morris report.

Although the committee studiously avoids taking sides, its tone of voice is consistently sympathetic to the universities. Thus it asks that when research council institutes are engaged on research that might also be carried on in universities, the research council concerned should "undertake a review" of its arrangements.

The general theme of the report is that "we are firmly convinced that there should be a much closer integration of the research institutes' and the universities' research activities". Sometimes, it seems, "suspicion and jealousy" impede relationships and, thinking positively, the committee asks the research councils to take the lead in improving relationships with the academic research community.

Other housekeeping points for the councils include the following:

- Research councils should review the work of their establishments at least once in

four years and make the reports of these investigations more widely available.

- In the interests of flexibility, councils should employ more staff on short-term contracts.

- Some research institutes should be made to compete for a part of their funds "on a similar basis to the universities".

- Institutes set up to provide a service to the universities should always (not just sometimes) have users' committees.

- The work of the same establishments should be reviewed at regular intervals, among other things to avoid the temptation that their own staffs will consume the services they are intended to provide for others.

There is some full-throated applause for those councils (chiefly the Medical, Agricultural and Natural Environment Research Councils) that support research units within universities, chiefly on the grounds that this is an effective way of stimulating a close interaction. And the universities do not get off scot-free — the committee complains of the inflexibility of their departmental structure and of their poor record in multidisciplinary research.

A note of something akin to idealism sounds through it all — the Morris committee urges that the research councils and their establishments should encourage a sense of community among people working in the same field at different institutions. And there is loud approval for the recent signs that polytechnics, the forgotten half of British higher education, may soon be treated as if they were capable of carrying out research. □

Curb on US political checks proposed

Washington

IN the wake of a series of revelations concerning the Reagan Administration's practice of screening candidates for federal scientific advisory committees for their political leanings, Senator Dale Bumpers (Democrat, Arkansas) has proposed legislation explicitly banning such practices. His bill (S. 1641) would prevent government officials from investigating the political affiliation of candidates for such committees or using such information in selecting committee members.

Reagan Administration officials have routinely submitted names of candidates for "political checks" by the Republican National Committee, where officials then report back whether they have voted in recent elections.

Bumpers's bill allows anyone to sue any federal agency to enforce the act and requires a committee to be discharged and reappointed from scratch if any of its members have been appointed — or any candidates denied appointment — in violation of the act.

Stephen Budiansky

*The support given by research councils for in-house and university research, ABRC, available free from Publications Despatch Centre, Dept of Education & Science, Honeypot Lane, Stanmore, Middlesex HA7 1AZ, UK.

Tarapur reactor

US spare parts deal held up

Washington

THE Reagan Administration's plan to supply India with spare parts for the troubled Tarapur nuclear reactor is under fire from two fronts, and the prospect now is for a long fight that may drag on for months. Last week, six environmental and anti-nuclear groups including the Union of Concerned Scientists filed a legal challenge with the Nuclear Regulatory Commission (NRC) seeking to block the sale. The challenge accuses India of violating the requirements of the Nuclear Non-Proliferation Act by refusing to accept full international inspections of its nuclear activities and by continuing its efforts to develop a nuclear explosives capability. India, which has refused to sign the non-proliferation treaty, exploded a nuclear device in 1974.

Earlier last month, 55 senators and congressmen — including Democratic presidential candidates Gary Hart, John Glenn and Alan Cranston — asked President Reagan to reject the State Department's recommendation, announced by Secretary of State George Shultz in New Delhi on 30 June, to release the spare parts, which have been held up since 1980.

The State Department is hoping to avoid a confrontation by persuading Italy or West Germany, which own reactors of a type similar to that at Tarapur, to supply the needed parts to India. It seems likely, however, that at least some of the parts are not available from these third parties and will have to come directly from the United States.

NRC will have to pass on those exports coming from the United States. In a briefing to congressional staff, NRC has already indicated that on the basis of intelligence information it has received on India's nuclear explosives programme, it will have to block the sale. The Nuclear Non-Proliferation Act requires that nations receiving nuclear assistance from the United States should not be "engaged in activities having direct significance for the manufacture or acquisition of nuclear explosive devices". According to congressional staff, the intelligence information includes satellite photographs of new test holes in the Rajasthan desert where India detonated a nuclear device in 1974.

An NRC finding against the sale would put the President in a sticky position. Although he can under the law order a waiver that would allow the sale to go ahead, he could only do so by first issuing a written determination that India was indeed in violation of the non-proliferation requirements — precisely the diplomatic embarrassment the administration has been trying to avoid.

The law also gives Congress 60 days to override the President's waiver, but this provision has almost certainly been nul-

lified by the Supreme Court's recent decisions that congressional vetoes are unconstitutional.

The problems with the Tarapur reactor stem from defective fuel elements made by the Indians from low-enrichment uranium supplied by the United States. As the cladding of the fuel has deteriorated, the coolant has become contaminated; leaking seals, flanges and pumps are allowing significant amounts of radioactivity to escape. In announcing the administration's intention to supply these replacement parts, Shultz stressed the need to prevent further exposure to radioactivity of plant workers and nearby residents.

The Indian press has described "bucket brigades" of workers hired to dash into contaminated regions of the plant, twist a wrench a few turns and then dash out before accumulating an excessive radiation dose. Although the Indian Embassy in Washington discounts these reports, NRC staff is said to have reported that Tarapur workers were given iodine pills, a preventive measure that blocks uptake of radio-nuclides by the thyroid gland.

Opponents of the parts sale argue that the situation is probably not as critical as the administration suggests. A General Electric engineer who visited the plant last month reported that both of the 210-MW reactors at Tarapur were operating and that no exceptional problems were

apparent. But opponents of the sale also contend that the "marginal improvement" that would be gained by supplying the parts would not make the plant safe. NRC officials have said privately that a plant of the Tarapur design and condition would not be granted an operating licence within the United States; safety features ordered in the wake of the Browns Ferry plant fire and the Three Mile Island accident have not been incorporated into Tarapur.

The main point of denying India the parts would be to maintain pressure on India to drop its nuclear explosives programme and to accept full international safeguards on its nuclear activities. Tarapur itself has not so far been involved in any of the explosives work. But concern is mounting that this situation may change; although a reprocessing facility at Tarapur is currently under International Atomic Energy Agency inspection, India has refused to offer assurances that the inspections will continue after 1993, when the US-India joint agreement that established Tarapur expires. The United States has refused to give its permission for reprocessing of spent fuel from the Tarapur reactors; India maintains that it does not need US permission. Earlier this year, the director of India's Atomic Energy Commission was quoted as saying that reprocessing of Tarapur spent fuel — which contains about one tonne of plutonium — would commence at the end of this year; he subsequently disclaimed the statement, saying he had been misquoted.

Stephen Budiansky

French nuclear tests

Tazieff report urges change

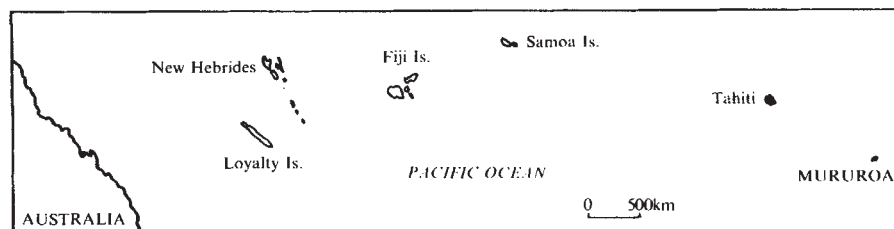
CHARLES Hernu, French minister of defence, has promised to act on a report that criticizes the safety and environmental monitoring of French nuclear tests in the Pacific. Remarkably, this report is French, and official, compiled by a seven-man team of scientists headed by geologist Dr Haroun Tazieff. But Tazieff is also director of the "centre for the study of major natural risks", supported by the Prime Minister's office, and some are dismissing the report being as biased too much towards the government point of view.

In fact, the report does indeed play down most of the risks of the tests — and gives no figures, using terms such as "feeble" and "innocuous" radioactivity — but it does stress wide areas of ignorance and hints at

poorly organized monitoring. One of the main releases from the atoll, for example, appears to have come from a storm in March 1981, which tore up part of a dump containing plutonium waste from aerial tests. (These tests ended in 1974, according to ministry of defence sources.)

But according to one member of the Tazieff team, who, like each other member, appends a personal view to the agreed report, the monitoring system "ignored the quantities introduced into the environment, their solubility and possible dispersal". Some "10–20 kg" of plutonium may have been mobilized, another scientist estimates, in one of the few figures in the report.

Another area of uncertainty concerns the sediments around the atoll, the report



says. Bathymetry has been meagre, and yet there is a risk of sudden slippage of sediment caused by the shock of an explosion which could launch an asymmetric sea-wave kilometres long and of about one minute period — something between ocean swell and a tidal wave. Without detailed knowledge of the sediment form around the island "it is extremely difficult to calculate the degree of stability [of the sediments] and hence to calculate the level of risk". If launched, the wave could swamp Mururoa, which rises only two metres above sea level. Warning systems and escape platforms for Mururoa personnel should be strengthened, the report says.

As for general radioactive emissions from the aerial tests and the later tests underground, these are "feeble compared with other regions" and with respect to natural radioactivity, the report says. However, the Tazieff team was able to make few measurements of its own, and in fact criticizes the absence of French monitoring stations on surrounding islands, thus forcing critics to rely on measurements by foreigners.

The Tazieff team had itself to rely on data provided by the local monitoring groups, which were divided into three independent — and warring — teams (concerned with physical, biological and medical measurements). And the report of one of them (physical), which had estimated that plutonium-239 emissions from Mururoa were "a little less" than those of Cap de la Hague, the French reprocessing plant near Cherbourg, "did not respect the usual rules of scientific publications: absence of certain numerical results, mean values without errors, descriptions of sampling without indication of date, incomplete references".

The Tazieff group concludes that research on and around Mururoa must be strengthened. Releases of radioactivity are unlikely "in the short term" — from individual underground tests — but there is an unknown risk of long-term releases into the ocean if test chambers are connected with the sea through cracks in the coral basement of the atoll.

"The absence of such information disarms defenders of the French nuclear test programme", the report says. There should be official announcements of tests (frequently the news of a test first comes from seismic monitoring in New Zealand or Australia) and "the publication of unattackable scientific documents" that would contain "all measurements not directly relevant to defence secrets". This action, says the report, "would considerably improve the psychological climate" surrounding French testing.

"Safety has always been our major concern", replied the ministry. "But it is always possible to do better". The recommendations in the Tazieff report "will certainly permit us to make progress in this field."

Robert Walgate

UK rf exposure

New standards unrestrictive

MANUFACTURERS of microwave ovens in Britain will be relieved to learn that proposed new standards of human exposure to high-frequency and microwave radiation are unlikely to affect their operations. But some industrial operations, such as those in which high-frequency radiation is used for heat-sealing plastic envelopes, will be affected and may even become "impracticable".

These are among the conclusions of a survey carried out by the National Radiological Protection Board (NRPB), the source of the advice on which draft standards for the protection of people from the effects of non-nuclear radiation were published in Britain last year. S.G. Allen and F. Harlen of the board's staff have now assembled data from a variety of sources that suggest that television transmitters are not the ogres that they have sometimes been represented, but that some fixed radar installations (civil as well as military) could be restricted by a tightening of standards beyond those now proposed.

The British standards (like the equivalent US standards) start from the proposition that the rate of heat dissipation within the human body caused by radiation should not exceed 0.4 W kg^{-1} on the average or 4.0 W kg^{-1} in the most dissipative cm^3 of tissue. The consequence is a frequency-dependent limit on power intensity amounting to 10 W m^{-2} between 30 and 100 MHz, inversely proportional to the square of the frequency below 30 MHz and proportional to the frequency itself between 100 MHz and 1 GHz.

The report* embodying the results of the surveys that have been carried out says that while there may be overt damage to body tissues if heat dissipation causes temperatures to increase, stress of the body's thermoregulatory apparatus by the absorption of radio energy may also be damaging in unknown ways.

The report also gives some credence to reports in the British popular press in recent years that some people are able to recognize their presence in or out of radar beams by means of sounds heard within their heads. The cause of the phenomenon appears to be the movement of the various pieces of a person's skull as a consequence of *in situ* heating by radio absorption, and tends to be heard as a series of clicks in a pulsed radar beam. The NRPB report estimates that the new standard should ensure that the noises are not audible except perhaps in a rotating pulsed beam. (For a review, see James C. Lin, *Proc. IEEE* 68, 67-73; 1980.)

For the rest, there is some concern about the use of portable radio transmitters, with some evidence that a driver sitting too close to his antenna might be on the threshold of

*Sources of exposure to radiofrequency and microwave radiations in the UK, NRPB-R144.

safe exposure. Industrial microwave plants, which the report says are often "effectively unshielded", are likely to be "significantly" affected by the introduction of the new standards. This opinion is based on Swedish and Finnish measurements of field intensities around radio-frequency equipment. And it seems that physiotherapists administering diathermy treatment may also have to be restricted to keep them within the now proposed limits of exposure. □

Spacelab delay

THE launch of Spacelab, the joint scientific mission of the National Aeronautics and Space Administration (NASA) and the European Space Agency, will be delayed by a month, NASA announced late last week. The new launch date is 28 October.

The delay stems from continuing uncertainty about the condition of the Tracking and Data Relay Satellite (TDRS), which was placed into the wrong orbit when the shuttle Inertial Upper Stage malfunctioned during deployment from the shuttle last April. NASA scientists had been working since then to nudge TDRS into its correct geosynchronous orbit, which was finally attained last month.



Spacelab's pressurized module

NASA officials decided that the originally scheduled launch on 30 September of Spacelab, which will be carried aboard shuttle flight number nine, would not allow enough time for completion of ground tests of the TDRS communication systems. The acid test of TDRS will come from the in-orbit testing of the satellite's systems by the crew of shuttle flight eight, which will be launched on 30 August, ten days later than originally planned.

If TDRS turns out to have been damaged during the deployment malfunction, Spacelab will probably be delayed by many months; the second TDRS is not scheduled for deployment until March 1984, and even that date is contingent upon successful debugging of the Inertial Upper Stage in time.

Stephen Budiansky

UK accelerators

Texas jumps next big gun

Washington

TEXANS, never shy when it comes to thinking big or spending big, have made an early bid for the massive colliding-beam proton accelerator recently endorsed by a federal advisory panel (see *Nature* 14 July, p. 105). Undeterred by the fact that the project has not yet even been submitted for administration or congressional approval or by the absence of even a conceptual design for the accelerator, Texas has floated a proposal to pay for the site and the tunnel construction if it is located in the state.

Peter McIntyre, a physicist at Texas A&M University and the moving force behind the scheme, has enlisted the support of four Texas universities (besides his own, the University of Texas at Austin, Rice University and the University of Houston) and the governor of the state, Democrat Mark White. In a letter to the Department of Energy, White expressed his strong support for the proposal and pointed out the benefits for industrial growth in his state. An aide to the governor said that the state revenues, private sources including the state's oil companies and high-tech industries, and the endowment incomes of Texas A&M and University of Texas (which derive in substantial part from oil and gas leases) could be tapped to raise the \$250 million needed.

The host site for accelerators is usually given free, but construction costs are as a

high cost will rule out construction of any other new machines elsewhere, will ensure that the institution that gets it is the major particle physics research centre for some time to come.

"With the governor behind us and with the university administration behind us, we think our prospects are good", McIntyre said. He envisages using the available technology of relatively low-field superferic magnets (perhaps 2 tesla) in a very large ring (perhaps 100 miles in circumference) — the original "desertron" concept (or, as it is known in Texas, the "Texatron"). A workshop held last spring at Cornell University examined the possibility of using a smaller ring with higher field superconducting magnets; but McIntyre, citing the research and development needed to build such magnets and mass-produce them, is scornful of this approach: "It's always possible to convince yourself that

you can do a better job if you give yourself five years of R&D. That's the road to failure."

To some non-Texans, all of this seems premature. Burton Richter of Stanford University said that the Department of Energy has not even figured out how it will manage such a vast project. Richter said that before a proposal can be submitted to Congress, there has to be a research and development programme, criteria for selecting a site, a conceptual design and a realistic cost estimate.

James Leiss, who directs the Department of Energy's high-energy and nuclear physics programmes, expressed similar concern about getting too far ahead of the administration and Congress. "I think it's going to be three years in the R&D program before we have a site-selection criteria document written", he said. "I think it's much too early." And he added that Texas is not the only state that is interested. "It's obvious it's going to be an exciting facility that will have a lot of economic advantages."

Stephen Budiansky

Radiation workers

Health bias to recruitment?

TRUE to its policy in recent years of releasing information on employees' radiation and mortality, British Nuclear Fuels Limited (BNFL), the publicly-owned monopoly for reprocessing and fabricating nuclear fuel, has published a progress report on its epidemiological study on the subject. The results show that employees and ex-employees at BNFL's Sellafield site (formerly called Windscale) have a significantly lower mortality rate than the population of England and Wales as a whole. The survey covered both cancer deaths and deaths from other causes. The company expects that the survey will be of particular public interest because of the attention recently paid to the incidence of cancer in Cumbria where Sellafield is sited.

The new data, for the Sellafield site, extend earlier results to include deaths up to the end of 1980, and cover some 11,500 employees and ex-employees who worked at Sellafield for periods that began before the end of 1975. The study population is divided into serving employees, those who have retired and other ex-employees. Observed/expected mortality ratios do not differ significantly between cancer and other deaths, and mortality is no higher among those classed as "radiation workers" than among others. Nor is there any significant increase in the rate of leukaemia, thyroid cancer and multiple myeloma.

The significant deficit of total deaths, both in the "cancer" and "non-cancer" categories, is most probably explained by a phenomenon known as the "healthy worker effect", whereby those presenting themselves for employment are likely to be more healthy, in general terms, than the

population as a whole. This effect is expected to be smaller for cancer deaths, although the data show no such reduction, possibly because of the ban on smoking enforced on radiation workers.

The observed number of deaths from causes other than cancer among those who had retired is, however, significantly

Sellafield mortality, 1948-1980

		Radiation workers	All workers
Bone marrow, bone and thyroid cancers	Serving	6 (4.4)	6 (5.2)
	Retired	2 (1.5)	2 (2.1)
	Other	2 (3.6)	2 (5.4)
	Total	10 (9.6)	10 (12.7)
Multiple myeloma	Serving	3 (1.3)	4 (1.5)
	Retired	0 (0.6)	0 (0.8)
	Other	0 (1.1)	0 (1.7)
	Total	3 (3.1)	4 (4.1)

Observed and (in parentheses) expected numbers of deaths among past and present employees from bone marrow, bone and thyroid cancers and multiple myeloma. From *Further Report on the BNFL Radiation-Mortality Study*, by E.A. Clough. Copies available from British Nuclear Fuels Limited.

greater than for the population as a whole, and cancer deaths in this group border on a significant excess. The healthy worker effect cannot explain these excess deaths. One possibility is that those retiring include some who retired early because of sickness. The results released so far do not permit an assessment of the extent to which the "healthy worker effect" is responsible for lowering total mortality rates. One epidemiologist suggested earlier this week that the company might look for an effect due to total time of employment, or make a comparison with other industries not involving radiation.

Tim Beardsley



rule borne totally by the federal government. Texas's offer to carry the construction costs of the tunnel is something of a surprise. McIntyre said that proprietary research he is pursuing on tunnelling technology could substantially reduce the cost.

Texas A&M and the University of Texas have of late adopted an aggressive policy to build up their physics departments. Last year, the University of Texas scored a coup by luring Harvard physicist Steven Weinberg, apparently by offering him a particularly high salary; Texas A&M has been pursuing Sheldon Glashow, reportedly with an even heftier salary offer. The new accelerator, both because of its advanced physics potential and because its

European nuclear fusion

Torus takes summer break

EUROPE's collaborative experiment in nuclear fusion, the Joint European Torus (JET), produced "real plasma" for the first time last week — and was then shut down for August. The machine has been operating since 24 June at the site alongside the Culham Laboratory of the UK Atomic Energy Authority in Oxfordshire, England. Last week, the machine produced a current of 0.6 MA in the D-shaped torus as a pulse lasting for a quarter of a second.

This, however, is rather less than the 1 MA current that had been hoped for. "One month was just too short", one JET physicist said last week. Now the whole experiment will have to wait while the electricity utility, the South-Western Electricity Board, upgrades a circuit-breaker on the power line supplying Culham. The board says that the problem has not been caused by JET but by the need to modify throughout the United Kingdom all circuit breakers of the type used for linking two power stations to one customer.

The August shut-down had been planned for some time, but it is frustrating for the JET team. Most of the past month has been spent battling with impurities, the high-atomic number ions derived largely from the walls, which radiate energy and which thus reduce the conductance of the plasma. The walls had been baked at 200–300°C and subjected to glow-discharge for 250 hours, and it seemed that the machine was "cleaning up well" just before shutdown.

Earlier, a number of technical problems had delayed the operation of JET. Heating coils for the instrument ports in the vacuum chamber were delivered late, with the result that most runs have been carried out with cold ports, which have invited wayward condensation of impurities. Nevertheless there is confidence that JET will reach 4–5 MA in a few months' time, perhaps by the end of the year.

The goals are now to finish cleaning up JET when it comes on line again on 12 September and to investigate the control of the unique JET D-shaped plasmas. (The ring is like a ring-doughnut whose cross-section is a D.) Then successive levels of diagnostics — plasma measurement systems — will be tested, relevant software developed and the functioning of the plasma limiters (which stop the high-temperature plasma hitting and burning the walls) checked. Fuel injection will be investigated.

By mid-1984, JET should be ready to move out of this "phase I" into the 18-month "phase II", where successive levels of external heating of the plasma up to 5 MW thermal, using neutral beams and radio-frequency heating, will be tried. This will be a critical time: the neutral beam heating "pushes the available technology to the limit" while the antennae needed to

radiate the radio-frequency energy into the plasma "are an area of uncertainty", according to a JET physicist. The extra heating is believed to be essential to push a deuterium-tritium plasma up to the 10 keV temperatures (and the $2-4 \times 10^{20} \text{ m}^{-3}$ density times confinement time) needed for a self-sustaining fusion reaction.

Hungarian environment

Belated help for teachers

A NEW subject — environmental education and awareness — will make its appearance in the syllabuses of Hungarian teachers' training colleges in September. The course — which will include a compulsory examination — is part of a major drive to heighten environmental awareness in the whole Hungarian nation. According to the National Authority for Environmental Protection and Nature Conservation — Hungary's supreme coordinating body on environmental matters — considerable progress has already been made but a closer look at the situation suggests that much remains to be done.

To begin with, it is somewhat surprising that only now is a teachers' course being introduced, since environmental education

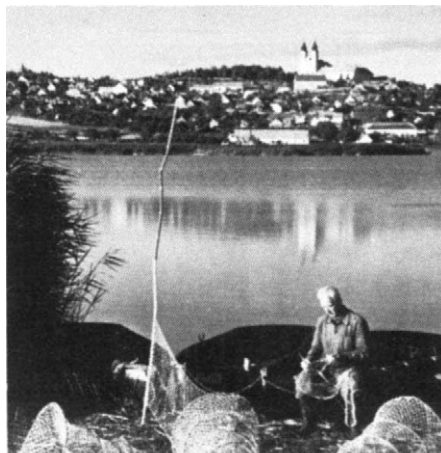
Phase III (another 2 years, taking the project to the end of 1987) should see more heating (up to 19 MW), and a decision on whether D-T ignition conditions are likely to be reached in the plasma core, and thus whether to proceed to D-T operation. Phase IV in 1988 would see D-T operation and ignition, leading to alpha-particle production and heating and such neutron activation of the apparatus that the experiment would finally be abandoned — as, of course, a success.

Robert Walgate

ten of public awareness of the environmental issues. Hungary's classic environmental problem is the eutrophication of Lake Balaton, the major water resort of Central Europe, which caused considerable concern in the mid-1970s and necessitated careful monitoring of fertilizers in the surrounding counties of Zala, Veszprem and Somogy. With fertilizers, by a logical association of ideas, went plant protection chemicals, the more potent types of which are not permitted south of Highway 61 on the northern shore of the lake.

Less obvious pollutants, such as acid rain, or the emission of hydrogen fluoride from an aluminium works, still attract little attention. Last month, Dr Bruno Straub, the government's leading adviser on environmental matters, appeared on a special television "write-in" programme, answering viewers' letters on environmental topics. Virtually all the questions dealt with the more apparent forms of pollution. If it is invisible and does not smell, it seems, the Hungarian public does not easily recognize the menace.

Vera Rich



Lake Balaton — under threat

was introduced in primary and secondary schools by the syllabus reforms of 1972. In 1974, a special Commission on Environmental Education was established, while in 1976, the Act on Protection of the Human Environment stressed that citizens should be made aware of the tasks and requirements of environmental protection by means of schooling, training, public education and information campaigns.

Moreover, in spite of the 1972 reforms, one important sector of secondary education still has not been fully brought into the environmental network — the vocational training schools for the 14–18 year olds, which are the main source of future skilled workers in industry and agriculture.

There is some doubt, too, about the ex-

Estonian on trial

AN Estonian physicist, Dr Johannes Hirt, who in 1962 was awarded a Lenin Prize for research in construction technology, could face a death sentence for "embezzling state property on an especially large scale". Details have just reached the West of the trial which opened on 21 March and which promises to be one of the longest-running events in Soviet Estonian legal history.

Dr Hirt, who was managing director of the "Desintegrator" Construction Technology Bureau in Tallinn, is charged, together with five of his employees, with having defrauded the state of 1.2 million rubles. On another count, he and six employees are charged with giving bribes. On all 18 counts of the indictment, Dr Hirt protests his innocence and maintains that the whole trial is based on "envy and intrigue". He does, however, acknowledge that he is the author of several *samizdat* texts confiscated during a police search of his home. The indictment on this charge, however, has been separated from the trial now under way.

Vera Rich

Laboratory animals

The case for the defence

SCIENTISTS who perform experiments on living animals should make a greater effort to demonstrate to the public the benefits being gained, in the view of Dr Walter Bodmer, director of research at the Imperial Cancer Research Fund.

Dr Bodmer was speaking last week at a gathering of scientists and others with interests in animal research, brought together by the Association of British Science Writers to present "the case for vivisection" to journalists.

The organizers had felt that scientists were in danger of losing ground by default in the battle for public opinion on the use of animals in research, due to their reluctance to expose themselves to hostile criticism and, deplorably, possible abuse. It was ironic therefore that the Department of Health and Social Security declined to send a representative to explain the department's policy on animal tests for the safety of cosmetic products, and one speaker who did attend could be persuaded to attend only after being shown a list of the strictly-invitation-only guest list. In the event, the largely uncontroversial addresses were mainly restricted to lists of important research and safety testing that could not proceed without experiments on living animals (monoclonal antibodies, immunosuppressive drugs, oncogenes and

cancer, quality control of artificial hormones and so on).

Mr Bryan Cassidy, of the Toiletry and Perfumery Association Ltd, emphasized that (contrary to the impression generated by some antivivisection groups) the number of animals used in testing cosmetic products is a very small proportion of the total number of animal tests in Britain. The tests that are made, he claimed, are mainly on products with some medicinal effect (such as anti-dandruff shampoos and fluoride toothpaste). The popular image of rabbits being force-fed decorative products such as lipstick is a travesty, because such products are based on long-used ingredients well known to be safe.

However, Mr Cassidy did complain that European Community regulations are obliging manufacturers to produce full toxicological dossiers on well known substances. The only other speaker prepared to step beyond the justification based on putting human lives and suffering before those of other animals was Professor Ian Steele Russell, Professor of Neuroscience at University College, London, who bravely made the point that research must continue even if only for intellectual curiosity, since the practical implications of basic research could often cannot be foreseen.

Tim Beardsley

UK universities

Part reprieve this time

FEARS that public support for British universities and for research would be cut in the latest round of government economies have not materialized. On the eve of the long parliamentary recess, the Secretary of State for Education and Science, Sir Keith Joseph, announced last week that grants to the universities and the research councils will stay as they were. But the department intends to claw back from the funds already allocated to the University Grants Committee £23.5 million of the cost of university "restructuring" on the grounds that the original sum was "larger than is now necessary".

The government's plan to reduce public expenditure in the financial year now one-third gone by a total of £1,000 million was announced at the beginning of last month. (Half of the total will come from the disposal of part of the government's residual shareholding in British Petroleum Ltd.) These immediate economies are distinct from the exercise also now under way to reduce projected public expenditure by £5,000 million in future financial years.

Last week's statement says explicitly that there will no reduction of the grants to individual universities which have already been settled and no reduction of the budgets of the research councils already

agreed for this financial year. Moreover, student maintenance grants, which are paid by local authorities and whose total amount cannot legally be limited, will remain at the level agreed for the coming academic year — an increase of 4 per cent.

For the rest, the department hopes to save modest sums by more efficient administration, and to save £5 million on what it had planned to spend on grants to higher educational institutions that it supports directly.

The most worrying aspect of last week's decision is that the funds set aside for university reorganization have been substantially reduced. While it may be that the actuarial cost of the premature retirement of academics may be less than originally foreseen (and the actual cost will not be known until September 1984), it is also clear that the account for helping university institutions to merge or to relocate themselves has not yet been drawn up.

These developments will be mildly reassuring for those who have been relying on the Prime Minister's promise in 1980 that the science budget would be "protected". Whether the promise will survive the next and larger round of spending cuts will not be apparent until towards the end of the year. □

French energy

Electricity glut the danger

THE French Government has decided to reduce its ordering of nuclear power reactors by a third, to two a year, this year and next, compared with a planned three. This is closely in line with the recommendations of the Commission du Plan (*Nature* 26 May, p.273), which by estimating future French energy needs — falling because of the recession — had calculated that France could get by with existing reactors until 1987. To save the nuclear industry, which had claimed that to stop ordering would kill it off and throw thousands out of work, the commission then recommended the order of two reactors this year, then one a year to 1990. The government will in fact order two this year, two next and "one or two" in 1985.

In the meantime, Electricité de France, the national utility, will be encouraged to go in for some aggressive marketing, to try and push up the industrial use of electricity. This, however, must imply a less bright future for other forms of energy, notably coal — a highly political issue given the large numbers employed in the coal industry in Prime Minister Pierre Mauroy's northern constituency — and gas, which France is committed to buy from the trans-Siberian pipeline and also from Algeria. Far from an energy shortage, France is heading for an expensive energy surplus.

What effect this will have on French industry in, say, ten years, is difficult to determine. According to Jean-Pierre Brunet, president of the Compagnie General d'Electricite, which sells all kinds of electrical equipment, the task is to find uses in 1990 for an extra 50 TWh of electricity. Brunet claims in an article in the French newspaper *Le Monde* that French homes and industry are electrically under-equipped. "And France could export electricity", he says.

Brunet identifies four areas in industry that could profit from surplus electricity: one is where the electrical energy is used to raise the temperature of low-grade heat, as in heat pumps or recompressing low-pressure steam from turbines, the electrical power thus benefiting from a thermodynamic "amplification". The second is where special properties of electricity can be used — Brunet mentions polymerization, vulcanization, sterilization, melting and inductive heating. Then there are "new activities" such as laser cutting and soldering of materials; and finally, dual use in heating, where traditional sources are used during peak hours and electricity used off-peak. But whether such ideas will be inventive enough to raise French electricity consumption by more than an eighth over the 380 TWh per year in 1990 foreseen by the Commission du Plan is doubtful.

Robert Walgate

European accelerators

Bonn-Hamburg e-p collisions

A £300-million West German accelerator seems very close to final approval by the federal government. The planned machine, called HERA, is designed to penetrate the intimate structure of the proton by colliding electron and proton beams, something that has never before been achieved.

The science minister, Dr Heinz Riesenhuber, had indicated a few months ago that HERA could go ahead provided foreign laboratories contributed some 20–30 per cent of the cost. Now, a contribution at that level seems guaranteed by a clutch of letters from laboratory directors in Italy, France, the Netherlands and Canada to Professor Volker Soergel, director of the Hamburg electron accelerator complex, DESY, where HERA will be built.

Britain, the Scandinavian countries and Japan, although invited, will not be participating directly. At Hamburg, they say that Britain will be particularly missed on account of the expertise in proton machines accumulated at the Rutherford

Appleton Laboratory near Didcot. DESY engineers are experienced only with electrons, which are different beasts from protons on account of their smaller mass and which consequently behave differently in accelerators. One Rutherford Appleton physicist, however, says "we will give all the help we can" when the Spallation Neutron Source, the laboratory's big capital project, is completed, but there will be no money.

Of the other possible partners, none has full approval yet but each thinks it will materialize. This will be sufficient to satisfy Dr Riesenhuber, a DESY spokesman believes. If he is correct, and formal approval comes in October, contractors could begin digging the HERA tunnel in January.

One strictly political obstacle may still get in the way. The Christian Democrat (right wing) federal government wants a written contract of special support for HERA from the Social Democrat (left wing) Hamburg state government. Hamburg has always strongly supported DESY, which has followed the prudent policy of contracting out to local industry as much as possible of its work. The inevitable political friction between Bonn and Hamburg may, however, hinder agreement.

According to normal regulations for large projects, Hamburg must put in 10 per cent of the HERA price tag (DM630 million (£160 million) in 1980 prices, or an estimated DM900–1,000 million in total by the planned completion date of 1990). But Bonn wants Hamburg to agree to pay more, and to give guarantees in case of cost overruns. In fact, Hamburg has already put up DM40 million for 1984, already more than 10 per cent of the annual cost during construction; but there is no written contract for future years.

The plan is that HERA will have a 6.3 km circumference tunnel, curving under the suburbs around the confined DESY site. The tunnel will house a conventional 30 GeV electron ring and a superconducting 820 GeV proton ring. Collisions between counter-rotating beams in the two rings should reveal detail of the way quarks and gluons combine to make protons, a study that is easy with point-like electrons. It is more difficult at the European Organization for Nuclear Research (CERN) near Geneva now, three-quark protons collide with three-quark antiprotons.

Italy, which will probably be the major partner of Germany, plans to build and pay for the 47 kilogauss superconducting bending magnets for the proton ring. France (at Saclay, near Paris) will develop quadrupoles (focusing magnets) and the Netherlands is working on correction coils. Canada, says a DESY spokesman, is also proving to be "a good partner".

Robert Walgate

Chinese technology

No imports for their own sake

CHINESE imports of high technology often fail to yield the intended economic benefit, an official of the Ministry of Foreign Economic Relations and Trade has been hinting. Speaking on Beijing Radio, Mr Chang Jiarui condemned the practice of "wanting to import technology and equipment simply for the sake of importing".

China's drive to import high technology began four years ago and was justified by a policy statement of Premier Zhao Ziyang which took the line that science and technology are common assets of mankind. This policy, said Mr Chang, has in some sectors allowed China, in two years or less, to bridge a gap of 20 to 30 years in product quality and technical performance.

These benefits, however, have not been as widespread as had been hoped, partly because of a lack of Chinese expertise in handling the new technology. The purpose of importing foreign technology, he reiterated, is "to increase production, create wealth, meet our needs and increase our capital stock".

With imports costing anything from a few thousand to several hundred million US dollars (as in the case of the Baoshan iron and steel complex in Shanghai), some proper system of control seems necessary. In a country the size of China, Mr Chang acknowledged, the delegation of power is essential. But this can operate properly only if there is a proper infrastructure of information, consulting and statistical offices as well as "non-governmental associations for the promotion of technological imports". This infrastructure seems to be almost entirely lacking in China at present.

Vera Rich

Celltech share sale

CELLTECH, the British biotechnology company founded three years ago, seems to have embarked on a familiar trail by selling shares to new investors. Baring Brothers, the London merchant bank, announced last week that it had privately sold a total of 1.8 million shares in the company at a price of £1.75 a share, buying some of them itself on behalf of its own managed funds. Existing shareholders in Celltech have at the same time taken up a further 1.8 million shares in the company.

The effect of this new share issue will be to give Celltech a further capital injection of £6.1 million, more than half the capital with which it began. Mr R.H.I. Perrott, Celltech's finance director, said earlier this week that the company had no immediate need of cash, having spent only half of its original capital so far.

But it is expected that substantial amounts of working capital would be needed to finance projects of various kinds now in the development stage, primarily in the pharmaceutical and diagnostic field. Celltech is not seeking to create a "war chest" of the kind accumulated by other companies, he said.

To the extent that the price paid for shares at a private placing is more than nominal, the transaction values Celltech at £27 million. The most significant change among the shareholders' list is that the British Technology Group, the public agency for technological innovation, has in the past month reduced its stake from 44 per cent to 28 per cent, by two successive sales of shares to other shareholders. □

Genentech-HP venture

THE Californian biotechnology company Genentech has signed up with Hewlett-Packard to develop a range of instruments for use in the industrial application of biotechnology. The new organization, whose creation was agreed on 20 July, will be managed by Hewlett-Packard, which will be the major owner of the new company and which expects to contribute \$10 to \$20 million in the next five or six years.

While gene synthesizers and similar equipment may provide the new company with something to make in the immediate future, both partners emphasized earlier this week that they have set their sights on the development of instruments that will have a natural place in scaled-up production processes.

To begin with, the new organization will consist of half a dozen Hewlett-Packard engineers with one Genentech scientist seconded as a consultant. □

US-European space cooperation

SIR — We were pleased to see (*Nature* 26 May, p.271) the interest and attention devoted to the work of the US National Academy of Sciences-European Science Foundation Joint Working Group on Cooperation in Planetary Exploration (JWG). We too believe that these discussions could lead to a new and important level of cooperation between Western Europe and the United States in scientific exploration of the Solar System — a cooperation which would be extremely beneficial to both sides.

Unfortunately, there were some inaccuracies in the article's rendition of the continuing discussions and in the premature anticipation of JWG's final conclusions. We would like to clarify several points.

(1) Contrary to the information given in the article, JWG has not reached a final decision on the cooperative planetary exploration missions that it will recommend to the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA), or on their sequence. While the missions described in your article are similar to the candidates which JWG has under consideration, it is a factual error to represent them in the aggregate as the concluding position of this advisory body.

(2) The US planetary core programme recommended by NASA's Solar System Exploration Committee is derived from a low-cost mission strategy made possible by innovative approaches to planetary spacecraft and payload instrumentation, and by a continuing programme of stable funding. The initial core programme of four missions primarily represents a US national effort; of course, where there are opportunities within the core programme for collaboration with international partners, these are endorsed and encouraged by JWG. Moreover, the US Space Science Board, the parent body of the US delegation to JWG, has taken the position that a programme of cooperative missions should be predicted on a separate, vigorous US national planetary programme. Under these circumstances, we were surprised to see your article assert that JWG would take three-quarters of the initial US national core programme and mandate it as an international programme.

(3) Finally, we are puzzled by the statement that "support" for this effort will be relatively easier to win from NASA than from ESA. The various elements of support — fiscal, political and scientific — will of necessity extend the decision-making process into various governmental areas. Although there are clear differences between the decision-making processes in the United States and in Western Europe, we have no reason to believe that support for a cooperative programme will be more or less easy to obtain from either of the two sides.

As you will appreciate, JWG is in the late stages of a complex and sensitive series of

discussions and negotiations on a matter which could lead to significant effects on the way some planetary exploration will be conducted into the far future, and which could have salutary effects in broader ways.

HUGO FECTIG

(Chairman, European Delegation)

Max-Planck-Institut für Kernphysik,
Heidelberg, FRG

EUGENE H. LEVY

(Chairman, US Delegation)

Department of Planetary Sciences,
University of Arizona,
Tucson, Arizona 85721, USA

Not the PhDs

SIR — The opinion page analysing scientific fraud in the United States (*Nature* 2 June, p.361) preserves an erroneous impression cultivated in many previous reviews elsewhere that PhD researchers are responsible for the recent crop of scandals. Some medical journal editors are reluctant to make sharp identifications but the plain fact is that MD researchers are overwhelmingly implicated. How is it possible to discuss causes and remedies without mentioning the professional species?

Regarding cause, the disciples of Socrates from their first day of laboratory instruction are taught the sanctity of data. By the time they finish a PhD programme, suspect candidates have largely been eliminated by a series of mentors. The disciples of Hippocrates, contrariwise, receive diminishing doses of basic science laboratory training and indoctrination as medical schools respond to budget pressures.

Regarding remedy, a PhD falsifying data loses his job and his career. The MD culprit loses his position also but by signing a promise to withdraw from research, he can enter a career of clinical practice and easily double his annual income. Different penalties result in different susceptibilities to temptation.

ROBERT E. KUTTNER

Veterans Administration
Medical Center,
North Chicago, Illinois 60064, USA

THE article did not put the blame on PhDs (or even younger MDs) but said "the indifference of many senior people . . . is more serious". Editor, *Nature*.

Suicide by nuclear war

SIR — The book review by Frank Barnaby (*Nature* 16 June, p.639) is ominous. When he says the nuclear arms race may be out of political control, what he means, and later implies, is that in a formally democratic society it is out of democratic control because it is in the hands of special interests. One of these is the scientific establishment itself. The record here is absolutely dismal. Except for the Federation

of American Scientists (an elite body), and more recently and quite surprisingly the American Medical Association, attempts to inject a note of reality into this issue have not only been effectively non-existent but subject to a censorship which, to say the least, is intellectually dishonest. In this respect science is behaving like the dishonest tailors in the fairy story "The emperor's new clothes".

Rather than put up with this I have resigned from societies (AAAS, AMS) which practise it. Anyone who values science as a cultural endeavour should ask themselves what is it contributing to the human enterprise? Otherwise we face the ultimate catastrophe predicted by Norbert Wiener, namely a computerized war in space which will make the Earth uninhabitable.

Fortunately we have, as a constituency to appeal to, the legendary "man in the street", so far left out of the picture, who is not as stupidly complacent as the armchair strategists seem to think. For him the idea of racial suicide is not appealing.

M. C. GOODALL

University of Alabama,
Birmingham, Alabama 35294, USA

Food for others

SIR — You comment on the \$20,000 million programme by the US Department of Agriculture to reduce food production in the United States and therefore to keep prices and profits up (*Nature* 23 June, p.645) and suggest that more money could be spent on genetic research and on other methods of reducing cost.

While this suggestion is certainly sensible and worthy of much consideration, another idea occurs to me. With over 800 million people destitute and in conditions of near or actual starvation (World Bank) throughout the world and with the numbers of destitute and hungry growing within the United States, where the 1930s style soup kitchens have now reopened, would it not also be a sensible idea to pay the farmers to produce more, not less, and to give the excess food to those who most need it? I would have thought \$20,000 million would have helped the world's destitute quite nicely!

Of course, this would mean that we would have to run the world as if people, not profits, mattered and that idea seems to be anathema to the administrations on both sides of the Atlantic. On the other hand, if the United States spent at least a little on feeding the poor Indian in Central America, the latter, with a full stomach and friendly northern neighbour, might be more disposed towards the Western rather than the Communist way of doing things, and President Reagan might actually succeed in "saving" Central America. But perhaps that is too much to hope for!

G. BARBER

City of London Polytechnic,
London E1, UK

The second millisecond pulsar

On the principle that no celestial object is unique, the discovery of another millisecond pulsar is overdue. That reported on page 417 has an unexpected interest.

WHEN the first millisecond pulsating star (PSR1937 + 214) was discovered more than a year ago, there were two contrasting views of what might happen next: either the 1.5-ms rotating neutron star was such an exceptional object, so near the edge of instability perhaps, that it would prove to be an isolated phenomenon, or the late discovery of this quickly spinning object was a sign that radioastronomers had previously neglected objects in this range, so that the first discovery would quickly be followed by several others. The truth is somewhere in between. Far from a flood of new discoveries of pulsars with periods in the millisecond range, there has been just one — that reported by Boriakoff *et al.* on page 417. Millisecond pulsars are plainly uncommon though not exceptional objects. The new pulsar has the convenience of being the principal member of a binary system, which explains why the report of its discovery comes complete with no fewer than half a dozen similar explanations of how it may have been formed.

Even so, it is worth recalling the enormity of what needs to be explained. All pulsars are neutron stars, and nobody has seriously quarrelled with Chandrasekhar's estimate that it needs a mass 1.3 times that of the Sun to make a neutron star rather than a degenerate dwarf in which atomic nuclei retain their identity. Most pulsars have periods measured in seconds, which is explicable only because of the compactness and tensile strength of neutron matter (derived from the strong nuclear forces holding them together), which enable these stars to survive the stresses of rapid rotation. As it happens, the rotational speed of a neutron star is limited not so much by its tensile strength as by the rate at which a rapidly spinning object would lose energy by means of gravitational waves caused by readjustments of its shape.

Speed apart, the first millisecond pulsar had two other exceptional properties. The rate of deceleration of the spin, measurable enough, is small compared with the decelerations of more conventional pulsars. But since the deceleration of the rotation of a pulsar is a consequence of the interaction of matter with its magnetic field, the strength of the magnetic field of this quickly spinning pulsar must also be small compared with that of more conventional pulsating stars. So far as can be told from the data now published, the new pulsar PSR1953 + 29, with a rotational period of 6.1 ms, shares the same unex-

pected properties — small deceleration and a relatively small magnetic field (estimated at about 10^{10} gauss).

The fact that the new object is a partner in a binary system is the starting point for the accompanying accounts of how it may have arisen. Indeed, the new pulsar is only the fourth to have been found within a binary system, much less than the proportion (more than a half) of ordinary stars included within binary systems. No doubt the explanation is that the occurrence of a supernova explosion within such a system will often send the residual neutron star scurrying away. The consequence, however, is that it has been possible to glean something of the mass of the system from half a period's observation of the Doppler shift of the radio pulsation. The conclusion that the orbit is virtually circular is only tentative. The conclusion that the pulsar is more massive than its companion is not compelled by the data but is likely to be true. The accompanying explanations agree that the companion is now a white dwarf with a mass about 0.3 that of the Sun.

The explanations of the origin of the 6.1-ms pulsar follow similar lines. The common notion is that the now unseen component of the binary system has within the comparatively recent past (ten million years?) gone through the giant phase in the normal evolution of a star. To begin with, the orbit was much smaller than at present. When most of the internal hydrogen had been converted into helium, the star became a giant, burning hydrogen only in its envelope which by then, in any case, would have reached beyond the gravitational equipotential point between the two stars. Material would then have been transferred from the companion star to the primary together with angular momentum. So, the argument goes, the end result would be that now observed — a rapidly pulsating neutron star whose mass is comparable with that of the Sun and a companion in a nearly circular orbit that has lost all but the helium core with which it embarked on its giant phase. It seems to be agreed that if the dwarf continues to burn hydrogen in its envelope, it should be two orders of magnitude more luminous than the Sun and thus just on the limit of detectability even at a distance of 3,500 parsecs.

By sheer good luck, Helfand *et al.* (see page 423) have important information on this point. The new pulsar happens to be within one minute of arc from another

from which X-ray emission was sought during the lifespan of the Einstein X-ray satellite, with the result that it has been possible to use data to set an upper limit of the X-ray emission from the new system. The upshot of the argument is that the most likely progenitor of the pulsar was not a neutron star as such but a white dwarf which, by accreting matter from its once-giant companion, exceeded the critical mass of 1.3 times the solar mass, pushing it over the threshold separating dwarfs from neutron stars.

Everybody seems to agree that whatever the original condition of the system, it must recently have gone through a phase in which it would have been recognizable as a low-mass X-ray binary of the kind described by Virginia Trimble in her recent field-guide to binary stars (*Nature* 303, 137; 1983). Ruderman and Shaham carry the argument a step further in their account of the origin of the apparently isolated 1.5-ms pulsar. They conclude that there is a high chance (determined by the precise details of the orbit and the relative masses of the two stars) that in a binary system consisting of a white dwarf primary and a secondary that has become a giant, tidal forces may so disrupt the secondary star when the mass transfer has gone some way that it will either be totally fragmented or reduced in size to an undetectable planet the size of, say, Saturn.

All this is good clean fun of the kind that keeps astrophysics entertained — and entertaining. The discovery of the 6.1-ms pulsar does, however, reinforce the opinion that has been growing for some time that there are two kinds of pulsars — those formed in supernova explosions (usually with modest rotational speeds) and those formed by the accretion of matter onto white dwarfs or other degenerate objects. Pulsars of the former class have relatively small rotational speeds (with periods measured in seconds), large velocities perpendicular to the galactic disk and strong magnetic fields. The second class, consisting of pulsars formed by accretion onto a degenerate star, are represented by the two millisecond pulsars so far found and possibly by some of those with smaller rotational speeds. Telling which are which will always be difficult at the borderline. Looking for X-ray binaries that may be nascent systems like PSR1953 + 29 should, however, be rewarding — Trimble explains that there is a host of candidates.

John Maddox

Third World development

The development of marginal lands in the tropics

from Gordon R. Conway, Ibrahim Manwan and David S. McCauley

THE past decade has seen great progress in the agricultural development of the irrigable lands of the tropics, particularly in South-East Asia. Today it is common for two crops of rice to be grown on such land with yields of 5 tons or more per hectare per crop. Attention is now turning to the development of those lands which, on various criteria, are more marginal. Following an earlier workshop¹ which examined the consequences of the intensification of better-quality land in Indonesia, a further meeting was held in January 1983 at Malang, East Java, under the auspices of the Centre for Agricultural Research Programming and the Ford Foundation, to consider the specific agricultural, ecological and socio-economic problems of developing marginal lands.

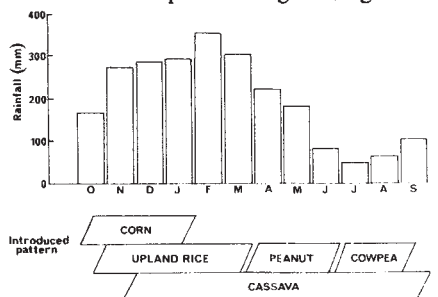
In Indonesia there are three major categories of marginal land: the tidal swamp-lands, primarily of Kalimantan and Sumatra (approximately 35 million ha); grasslands covered by *alang-alang* (*Imperata cylindrica*; about 15 million ha); and 'critical' uplands, mostly on Java and Bali, which are defined as lands suffering from severe degradation because of erosion (between 10 and 40 million ha).

In total these lands may cover as much as one-third of Indonesia's land surface, encompassing a wide range of ecological and socio-economic conditions. Thus whereas the development of the better-endowed lowlands has been achieved by disseminating widely adapted crop varieties and cropping techniques, the more diverse and severely constrained marginal lands require a more finely tuned approach. As a first step towards targeting the areas for research and development, the workshop participants felt strongly that the land should be zoned not only according to agroecological factors, but also in terms of socio-economic criteria. *Alang-alang* land, for example, should be characterized not only by climate, soil and topography but also by who owns or cultivates the land, their cultural and economic circumstance and, in particular, whether they view *alang-alang* as a weed to be eradicated or as an asset to be preserved.

Marginality arises partly from limiting factors, for example acid sulphate soils in the tidal swamps, low levels of soil nutrients in *alang-alang* land and steep slopes in the critical uplands; and partly from their inherent environmental instability. Tidal swamplands are subject to considerable seasonal and daily fluctuations in levels of water and salinity. *Alang-alang* lands experience frequent burning and, in some areas, severe

drought, while critical uplands suffer from periods of intense rainfall.

Such variability tends to be viewed as a constraint, but it can also provide opportunities for development. For example, the critical problem in the tidal swamps is that excessive drainage leads to destruction of the surface peat layers followed by acidification of the underlying soil and the production of toxic aluminium. Expensive engineering works



New cropping pattern for Indonesia's red-yellow podzolic soils³

which carefully control the water level are one solution, but in Indonesia areas for rice production have been successfully opened using simple communally operated gates which permit tidal flows to flush away the acids yet retain sufficient water to prevent oxidation.

A similarly simple and inexpensive solution is the use of burning to manage *alang-alang*. Although often an indicator of degraded and apparently abandoned land, anthropologists argue that in many cases *alang-alang* is a productive resource. In South Kalimantan *alang-alang* is an essential part of a rice-fallow system; on Sumbawa island it supports productive game hunting — the deer being attracted by the regrowth after burning — and in Bali it is an important cash crop, providing the materials for traditional thatched roofs which are coming back into favour.

Many of these traditional agricultural systems², although not highly productive, are sustainable, having evolved resilience to stress and perturbation. But under the demands of population pressure and economic necessity they may give way to less appropriate systems. The challenge is how to increase the productivity while retaining sustainability.

One answer is to blend new varieties and techniques with old patterns. For example, swamplands traditionally have been cultivated using photoperiod-sensitive rice varieties that are transplanted two or three times and mature over 8–10 months. New early-maturing photoperiod-insensitive varieties are now being sown on part of the

land to produce a quick first crop while the traditional varieties are still being transplanted. Sustainability is retained if drainage is not excessive and the level of organic matter is maintained.

A particularly productive and apparently sustainable cropping system has been designed^{3,4} for the red-yellow podzolic soils which occupy about 15 per cent of Indonesia's land area. Traditional cultivation consists of a mixture of crops followed by a fallow of *alang-alang*. The new system developed by the Central Research Institute for Food Crops comprises a more organized inter- and relay cropping, grown in a continuous cycle without a fallow (see the figure). About 1 ton per ha of burned limestone is applied initially, phosphorus is spread in the furrows, nitrogen and potassium are placed below and beside the seed, and all crop residues are returned as mulch. The benefits lie in the evenness of labour demand and the steady flow of produce and income. Five years of such continuous cropping have produced yields, in food calorie terms, of 12–25 tons per ha per year of paddy rice equivalent.

The development of sustainable systems for the critical uplands will be more difficult. The emphasis of government programmes has been on reforestation of both public and private lands and more recently on the construction of bench terraces and check dams. But such projects are generally costly and of limited immediate benefit to upland farmers.

An alternative strategy is to develop and extend traditional agro-forestry systems, such as the home and forest gardens^{5–9}. These are typically multistoried and highly diverse cultivation systems with perennial tree crops and a rotation of mixed annuals underneath. They provide the farmer with a steady flow of food, fibre, wood and cash crops, and because of the high degree of nutrient recycling and the completeness of the ground cover they also conserve the soil even on fairly steep slopes.

A 'Committee on the intensification and sustainability of Indonesian agriculture' has been formed to plan further workshops on specific target agroecosystems and to promote and seek funds for further policy and field research in this area. □

Gordon R. Conway is Professor of Environmental Technology at Imperial College, London SW7 1LU; Ibrahim Manwan is Director of the Centre for Agricultural Research Programming of the Indonesian Ministry of Agriculture; and David S. McCauley is a resource economist with the Ford Foundation, Jakarta, Indonesia.

1. Conway, G.R. & McCauley, D.S. *Nature* **302**, 288 (1983).
2. Terra, G.J.A. *Neth. J. agric. Sci.* **6**, 157 (1958).
3. McIntosh, J.L., Ismail, I.G., Effendi, S. & Sudjadi, M. *Proc. int. Symp. on Distribution, Characteristics and Utilisation of Problem Soils*, TARC, Tsukuba, Japan (19–20 Oct 1981).
4. McIntosh, J.L., Effendi, S. & Ismail, I.G. *Ind. agric. Res. Dev.* **2**, 13 (1980).
5. Nair, P.K.R. *Gartenbauwissenschaft* **42**, 145 (1977).
6. Stoler, A. *Bull. Ind. econ. Stud.* **14**, 85 (1978).
7. Terra, G.J.A. *Mal. J. trop. Geogr.* **1**, 33 (1953).
8. Wiersum, K.F. *Tropical Ecology and Development*, 515 (1980).
9. Wiersum, K.F. *Agrofor. Syst.* **1**, 53 (1982).

Nucleic acid synthesis

Pinyin tRNA

from Beverly Griffin and Tomas Lindahl

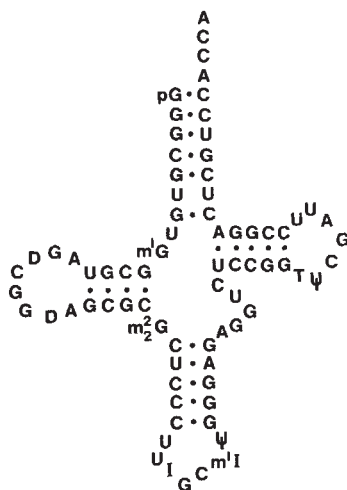
A LARGE-SCALE research project conducted in the People's Republic of China has recently been brought to a successful conclusion. In the May 1983 issue of *Scientia Sinica*, a series of four papers (written in English) describe the total synthesis of a biologically active tRNA^{1,2}. This is the first time such a molecule, indistinguishable from a natural isolate, has been synthesized.

The project was initiated as long ago as 1968, but progress seems to have been relatively slow in the difficult research climate existing during the Cultural Revolution. Much of the work has been carried out at the Shanghai Institute of Biochemistry of the Academia Sinica by a research group directed by Wang Debao (who published as T.P. Wang before the current *pinyin** transliteration system was generally adopted), but important contributions have also been made by other groups in Shanghai and Beijing. In fact, it is estimated by the Chinese workers that between 100 and 200 scientists have been involved in this project at various times, but this staggeringly large figure includes units such as the 'Shanghai No.2 Reagent Factory', which produces fine chemicals and precursors often readily available commercially to Western scientists.

While modified and improved chemical methods for DNA synthesis have made the production of oligodeoxyribonucleotides of defined sequence a widely used technique of molecular biology within the last few years, synthesis of related oligoribonucleotides is considerably more complicated because of the presence of reactive 2'-OH groups on the sugar residues. Even today, it would seem forbiddingly difficult to try to synthesize a RNA molecule the size of a tRNA exclusively by chemical procedures. Therefore, the strategy finally adopted by the Chinese workers involved the chemical synthesis of several small oligoribonucleotides, which were then enzymatically joined together in a predetermined fashion. An analogous approach was pioneered for DNA by Khorana's group³, in the first synthesis of an active gene.

The RNA molecule chosen for synthesis by the Chinese scientists was the alanine-tRNA of baker's yeast, one of the very few complete RNA sequences known in 1968. This molecule contains 76 base residues, of which as many as 9 are modified versions of the four major nucleotide subunits (see the figure). Before synthesis, therefore, the necessary modifications had to be made to the monomer nucleotides by specific chemical reactions such as catalytic

hydrogenation, methylation and deamination. One of the minor residues, pseudouridine, which is not readily obtainable by chemical means, was isolated from human urine and subsequently phosphorylated. The detailed procedures for the incorporation of such unusual, and in some cases labile, nucleotide derivatives into oligonucleotides still await publication.



Primary structure of alanine-tRNA from baker's yeast. The nucleotide sequence is that proposed by Holley *et al.*⁸ with two subsequent minor revisions⁹. The modified nucleotide residues are D, dihydrouridylic acid; ψ, pseudouridylic acid; I, inosinic acid; T, ribothymidylic acid; m¹I, 1-methylinosinic acid; m¹G, 1-methylguanylic acid; m²G, 2-dimethylguanylic acid.

The discovery of the phage T4-induced RNA ligase⁴ and the demonstration that this enzyme effectively joins small single-stranded oligonucleotides together with no sequence-specificity or template requirements⁵ have been crucial for the synthesis of the tRNA *in vitro*. The other key enzyme used in the work was polynucleotide kinase, which allowed for the phosphorylation of 5' ends with a ³²P radioactive marker. The latter enzyme normally also has an associated 3'-phosphatase activity, but the Chinese team used an additional mutant form of the enzyme that has only the kinase action. By ingenious use of these activities, it proved possible for the Chinese scientists to use 3'-phosphate residues as blocking groups during the ordered joining of oligonucleotides, subsequently removing the blocking moieties by enzymatic means.

Two of the four papers in *Scientia Sinica* describe the separate syntheses of the 5' and 3' halves of the tRNA^{Ala} molecule¹. In a preliminary study, it was shown by the Beijing group that half-molecules of tRNA^{Ala}, obtained by limited ribonuclease

T1 cleavage of the molecule purified from yeast, can be efficiently rejoined with RNA ligase⁶. As detailed in the other two current papers^{1,2}, the corresponding synthetic half-molecules were joined with RNA ligase, and the completed tRNA was shown to possess full amino-acid-acceptor activity, as well as transfer activity *in vitro* in a reticulocyte-derived protein-synthesis system. The synthetic molecules were also shown to migrate in the expected position on polyacrylamide gel electrophoresis in the presence of 7M urea. It is perhaps typical of the current preference for gel electrophoresis over centrifugal analysis that no attempt was made to show that the synthetic molecule had the 4S sedimentation coefficient characteristic of the native compact tRNA conformation.

The principal difference between this impressive synthesis and earlier efforts to fabricate tRNA *in vitro* is the inclusion of all the correct modified residues in the synthetic molecule. Ohtsuka *et al.*⁷ have previously reported the synthesis of an RNA molecule with a sequence related to that of *Escherichia coli* formylmethionine-tRNA, using an approach similar to that adopted in Shanghai. Although the work is very elegant, their final product was entirely composed of unmodified residues and showed little amino-acid-acceptor activity. Similarly, when the synthetic gene for an *E. coli* tyrosine-tRNA was transcribed *in vitro*, the transcript was inactive in an amino-acid-acceptor assay, although the gene could be shown to be active *in vivo*³. Taken together, these data dramatically demonstrate the crucial role of the modified nucleotides in conferring biological activity on tRNA molecules.

With the present achievements in synthesis, and the availability of the appropriate building blocks, the road now seems open for an unequivocal assignment of the relative roles of various base modifications for enzyme recognition or as key structural features in tRNA, since one after another of the altered residues can now be replaced by its unmodified counterpart in a synthetic molecule which can then be assayed for biological activity. This promising line of research is now being pursued by the Shanghai group. □

Beverly Griffin and Tomas Lindahl are with the Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX.

1. Wang Debao *et al.* *Scientia sin.* B26, 464, 482, 495 (1983).
2. Shen Qingxiang *et al.* *Scientia sin.* B26, 504 (1983).
3. Sekiya, T. *et al.* *J. biol. Chem.* 254, 5787, 5802 (1979).
4. Silber, R., Malathi, V.G. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* 69, 3009 (1972).
5. Walker, G.C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 72, 122 (1975).
6. Wang Gui-Hai *et al.* *Biochim. biophys. Acta* 652, 82 (1981).
7. Ohtsuka, E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 78, 5493 (1981).
8. Holley, R.W. *et al.* *Science* 174, 1462 (1965).
9. Penswick, J.R., Martin, R. & Dirheimer, G. *FEBS Lett.* 50, 28 (1975).

*Pinyin is the standard system for transcribing Chinese characters into Roman ones.

Stellar evolution

Are population II binary systems inherited or acquired?

from Virginia Trimble

ACCORDING to astronomical folklore¹⁻³, the oldest stars in our Galaxy have far less than their fair share of most kinds of binary system. These old stars, found both in globular clusters and scattered singly through the galactic halo, are collectively called population II, in contrast to younger, more metal-rich population I stars, such as the Sun, which are found in the galactic disc. Within the globular clusters, there are so far no convincing cases of either eclipsing binary systems²⁻⁵ or radial velocity variables (spectroscopic binary systems⁶⁻⁸) while the scattered halo stars include only a few spectroscopic binary systems⁹.

Two causes for the deficiency are possible: either binaries never formed (in which case we are learning something about how conditions of star formation have changed over the past 10 billion years) or the star pairs have somehow merged or broken apart (in which case we are learning about dynamical evolution of globular clusters and binary systems). Either would be interesting; and the recent identification of a new class of X-ray sources in globular clusters¹⁰ may enable us to decide what is going on.

Oddly, while globular clusters lack normal binary systems, they are oversupplied with strong X-ray sources by a factor of about 100 (ref. 11) and with cataclysmic variables by a factor of 10 or more^{2,3,12} relative to population I. And the best models of these X-ray sources and cataclysmic variables all invoke highly evolved binary systems with white-dwarf or neutron-star components¹³⁻¹⁵. What are

we to make of this? Several groups of theorists¹⁶⁻¹⁸ have concluded that the X-ray objects result from tidal capture of a main-sequence star by a previously single neutron star. Similar processes should clearly produce binary systems with other sorts of component star.

Hertz and Grindlay¹⁰ believe that their new X-ray sources are the white dwarf-main-sequence star pairs (cataclysmic variables) formed by such tidal captures. They report 14 new point-like sources, distributed among eight clusters (one to five per cluster), with luminosities of 10^{32} – $10^{34.5}$ erg s⁻¹ (0.5–4.5 keV), well separated from the $\geq 10^{36}$ erg s⁻¹ globular-cluster X-ray sources previously known and attributed to accretion in neutron-star binaries. The new source locations, some of which are well outside their cluster cores, indicate masses of ≤ 1.0 solar masses, appropriate for white dwarf-main-sequence pairs in this context. The distribution of source luminosities is like the bright end of

that for cataclysmic variables found in the galactic plane¹⁹; and extrapolation down to 10^{31} erg s⁻¹ implies about 20 white-dwarf binary systems per typical cluster. This, with calculated capture rates¹⁷, in turn implies about $10^{4.1}$ white dwarfs per cluster. These stars are too faint to observe directly in most clusters with current techniques, but models of globular cluster structure^{20,21} suggest that the total number is about right.

Hertz and Grindlay¹⁰ therefore conclude that the X-ray sources, bright and faint, and cataclysmic binary systems in globular clusters are the product not of normal binary-system evolution¹³, but of tidal captures. They agree that tidal capture to form main-sequence and giant pairs must also occur, but believe that the pairs so formed will be concentrated in crowded cluster cores sufficiently to have escaped detection as either eclipsers or radial velocity variables. I have previously argued²² that this would not be the case; but I may have been wrong. If so, then, evidently, globular clusters are not born with binary systems in place (which will require explanation in due course in any complete model of star formation) but acquire them late in life through tidal capture. □

Virginia Trimble is Visiting Professor of Astronomy, University of Maryland, College Park, Maryland 20742.

Protozoology

Gamete fusion in *Babesia*?

from F.E.G. Cox

THE babesias, common blood parasites of mammals, cause serious diseases such as Redwater Fever in domesticated animals and occasionally infect man, often with disastrous consequences. For many years it was believed that the life cycle in mammals consisted solely of repeated asexual binary fissions within the red cells and although circumstantial evidence suggested that a sexual stage was involved, more direct evidence of gamete fusion has only just been forthcoming¹.

Babesias are transmitted by ticks and it is within their guts that the search for sexual stages began. In 1906, Koch² recognized unusual forms which he called Strahlenkörper (ray bodies). These seemed to consist of a head, tail and arms and bore a superficial resemblance to the microgametes of other sporozoans. These stages were largely ignored until the electron microscope confirmed their existence and in recent years the life cycles postulated for *Babesia canis*³ in dogs and *B. bigemina*⁴ and *B. bovis*⁵ in cattle have included short-lived forms with spike-like projections as part of a sexual process. The various authors involved, however, have avoided the real issue by admitting an unwillingness to use the term 'microgamete' or by using such words as 'supposed gametes'.

The most recent evidence for the existence of sexual stages has come by a round-about route. *Babesia microti* is one of the commonest blood parasites of small mammals⁶ both in Europe and America but was largely disregarded until it was implicated as a sporadic cause of severe infections in man in the United States⁷. The tick vector involved was found to be a previously unrecognized species *Ixodes dammini*⁸ and the availability of this tick, together with hamsters infected with *B. microti* from a human source, opened up avenues of investigation hitherto unavailable. It is this model that Rudzinska and her colleagues have now used in an attempt to provide definitive proof that babesias have a sexual stage¹.

Larval ticks were fed on infected hamsters and after 60 hours of feeding, parasites in the tick gut were still in the red blood cells, although some of them had developed new organelles, including an electron-dense 'arrowhead' at the anterior end. By examining serial sections to avoid mistaking division for fusion, Rudzinska *et al.* found that the arrowhead forms emerged from their red cells to fuse with other emergent forms without arrowheads. There is no doubt that the arrowhead forms are the equivalent of Koch's

1. Batten, A.H. *Binary and Multiple Systems of Stars* (Pergamon, Oxford, 1973).
2. Webbink, R.F. *IAU Symp.* **88**, 398 (1980).
3. Trimble, V. *IAU Symp.* **85**, 259 (1980).
4. Belserene, E. & Liller, W. (in preparation).
5. Hogg, H.S. *Publ. David Dunlap Obs.* **3**, no. 6 (1973).
6. Gunn, J.E. & Griffin, R.F. *Astr. J.* **84**, 752 (1979).
7. Harris, W.E. & McClure, R.D. *Astrophys. J.* (in the press).
8. Latham, D. Personal communication.
9. Carney, B.W. & Peterson, R.C. *Astrophys. J.* **251**, 190 (1981).
10. Hertz, P. & Grindlay, J.E. *Astrophys. J. Lett.* **267**, L83 (1983).
11. Kaiz, J. *Nature* **253**, 698 (1975).
12. Margon, B., Downs, R.A. & Gunn, J.E. *Astrophys. J. Lett.* **247**, L89 (1981).
13. Trimble, V. *Nature* **303**, 137 (1983).
14. Lewin, W.H.G. & Joss, P.C. *Space Sci. Rev.* **28**, 3 (1981).
15. Paczyński, B. in *Cataclysmic Variables and Low-Mass X-ray Binaries* (Reidel, Dordrecht, in the press).
16. Fabian, A.C., Pringle, J.E. & Rees, M.J. *Mon. Not. R. astr. Soc.* **172**, 15p (1975).
17. Press, W.H. & Teukolsky, S.A. *Astrophys. J.* **213**, 183 (1977).
18. Lightman, A.P. & Grindlay, J.E. *Astrophys. J.* **262**, 145 (1982).
19. Cordova, F.A. & Mason, K.O. in *Accretion Driven Stellar X-ray Sources* (eds Lewin, W.H.G. & van den Heuvel, P.J.) 141 (Cambridge University Press, London, 1983).
20. Da Costa, G. & Freeman, K.C. *Astrophys. J.* **206**, 128 (1977).
21. Illingworth, G. & King, I.R. *Astrophys. J. Lett.* **218**, L109 (1977).
22. Trimble, V. *Mon. Not. R. astr. Soc.* **178**, 335 (1977).

Strahlenkörper and the putative sexual forms of other investigators.

The sequence of events in the sexual cycle is this. Merozoites enter red cells in the mammal host and most of these become trophozoites (feeding stages) and subsequently divide by binary fission. Some, however, become potential gametocytes which cannot be distinguished in the blood of the mammal but develop into gametes still within blood cells in the gut of the tick during its long phase of repletion. Several gametes may form in a single red cell and all or some of these may have arrowheads. The arrowhead forms emerge from the blood cell to fuse with a form that does not possess this characteristic structure, to become a zygote. During its feeding period the tick produces a temporary peritrophic membrane around the blood meal⁹ and the zygote uses its arrowhead to bore through this membrane to enter the epithelial cells of the gut where it begins a cycle of division. As it passes through the peritrophic membrane the arrowhead structure becomes markedly less electron-dense presumably because of the loss of enzymes required for penetration⁹.

This study¹ raises as many questions as it answers. It would be too easy to infer that the arrowhead forms are microgametes. As several of these occur in one cell this is possible, but the retention of a male structure like the arrowhead well into the zygote stage is unlikely. This could suggest that the arrowhead form is a microgamete but this again is unlikely because of the production of more than one in a single cell — assuming that they all develop from one merozoite, which seems to be the case. There is also as yet no evidence of nuclear fusion, without which it is not possible to say categorically that sexual reproduction involving fusion of gametes has taken place.

A study of the details of the reproduction of the babesias together with an investigation of what structures pass from the gametes to the zygote is a task for developmental biologists. At one time the need to use ticks would have made such a task nearly impossible but Strahlenkörper, now known to be sexual stages, do develop *in vitro* in cultures of tick cells exposed to blood infected with *B. bovis*⁵ or *B. canis*¹⁰. This should bring this kind of investigation within the scope of many cell biological laboratories. □

F.E.G. Cox is Professor of Zoology at King's College London, The Strand, London WC2R 2LS.

1. Rudzinska, M.A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 2966 (1983).
2. Koch, R. Z. *Hyg. Infektionskr.* **54**, 1 (1906).
3. Melhorn, H. *et al.* *J. Parasit.* **66**, 220 (1980).
4. Weber, G. & Friedhoff, K.T. Z. *Parasitenk.* **53**, 83 (1977).
5. Droleskey, R.E. *et al.* *Res. vet. Sci.* **34**, 249 (1983).
6. Cox, F.E.G. *Mammal Rev.* **1**, 1 (1970).
7. Ruebush, T.K. *et al.* *Am. J. trop. Med. Hyg.* **30**, 937 (1970).
8. Spielman, A. *et al.* *J. med. Ent.* **15**, 218 (1979).
9. Rudzinska, M.A. *et al.* *Cell Tissue Res.* **221**, 471 (1982).
10. Melhorn, H. *et al.* *Zentralbl. Bakteriell. Abt. I. Orig. A.* **250**, 248 (1981).

Virology

Neutralizing site of poliovirus

from F. Brown

DURING the past 30 years the structural analysis of many viruses has allowed them to be grouped according to their physico-chemical properties, thus providing a framework of classification which is valuable in taxonomy¹. The techniques that have been developed during the same period for the characterization and sequencing of nucleic acids and proteins, coupled with the dissection of viruses into biologically active fragments, now allow detailed chemical analysis of these relationships and have made it possible to define biologically active sites such as those involved in virus neutralization. These advances have led to the possibility of developing vaccines based on small synthetic peptides, and a step in this direction is reported for poliovirus in this issue of *Nature*².

Because of their relatively simple structure, the picornaviruses have proved particularly amenable to this approach. The members of this family of viruses, which includes such examples as the agents causing poliomyelitis, foot-and-mouth disease, the common cold and hepatitis A, are composed of one molecule of single-stranded RNA and 60 copies of each of four proteins. Early work on foot-and-mouth disease virus had shown that the major immunizing site(s) was located on only one of the capsid proteins, VP1³⁻⁶. More recent evidence shows that the immunodominant site on other picornaviruses, such as Mengo virus⁶, Cocksackie B3 virus⁷ and poliovirus⁸⁻¹⁰, is also confined to a single protein, although Wimmer's group has indicated that other sites are present on other capsid proteins of poliovirus¹¹.

An even more precise location of the immunizing site(s) of foot-and-mouth disease virus has been established from work by Strohmaier¹², Bittle¹³, Pfaff¹⁴ and their colleagues. The position of VP1 on the genome of the virus had been determined by Sangar *et al.*¹⁵ and the sequence of the RNA coding for it was known from molecular cloning experiments¹⁶. This allowed Strohmaier's group to identify the positions on VP1 of immunogenic fragments derived by enzymatic cleavage *in situ* or by cyanogen bromide cleavage of the isolated protein. From these data they predicted that regions 146–154 and 201–213 would contain antigenic sites. Bittle *et al.*¹³ based their prediction on the observation that whereas most of the amino acid sequence of VP1 from different serotypes of the virus is conserved¹⁶⁻¹⁸, there were hydrophilic regions of high variability, one of which corresponded fairly closely with the 146–154 sequence predicted by Stro-

maier's group. It was reasoned that the variation was due to the availability of these regions for reaction with neutralizing antibody, rendering them susceptible to selective pressure. The more recent prediction by Pfaff *et al.*¹⁴, made from the likely position of an α -helix on the surface of VP1, also corresponds closely with the sequence at positions 146–154. The Bittle and Pfaff groups followed up their predictions by demonstrating that organically synthesized peptides 141–160 and 144–159 stimulate the production of neutralizing antibody. The activity of this region has been confirmed by the demonstration that regions 130–157 and 141–160 biosynthesized in *Escherichia coli* cells will also elicit the formation of neutralizing antibody at a level sufficient to protect against infection¹⁹.

Schild and his colleagues in London and Leicester⁸ have adopted a different approach to predict the site involved in the neutralization of poliovirus. They selected antigenic mutant viruses for their resistance to individual monoclonal antibodies and analysed their RNAs by ribonuclease T1 mapping. This revealed that a specific oligonucleotide was absent from the resistant mutants, indicating changes in the viral RNA represented by this nucleotide that had generated a new antigenic configuration. Sequencing studies identified the oligonucleotide as a 16-base region 278–293 bases downstream from the start of VP1. The region corresponds to a tract of amino acids which is poorly conserved between serotypes 1 and 3 of the virus.

In an extension of this work², the same group now presents evidence that resis-

1. Matthews, R.E.F. *Intervirology*, **17** 1 (1982).
2. Evans, D.M., Minor, P.D., Schild, G.C. & Almond, J.W. *Nature* **304**, 459 (1983).
3. Wild, T.F., Burroughs, J.N. & Brown, F. *J. gen. Virol.* **4**, 313 (1969).
4. Laporte, J., Gosclaud, J., Wantyghem, J., Serge, B. & Rouze, P. *C. r. heb. Séance, Acad. Sci., Paris* **276**, 3399 (1973).
5. Bachrach, H.L., Moore, D.M., McKercher, P.D. & Polatnick, J. *Immunology* **115**, 1636 (1975).
6. Lund, G.A., Ziola, B.R., Salmi, A., & Scraba, D.G. *Virology* **78**, 35 (1977).
7. Beatrice, S.T., Katze, M.G., Zajac, B.A. & Crowell, R.L. *Virology* **104**, 426 (1980).
8. Minor, P. *et al.* *Nature* **301**, 674 (1983).
9. Blondel, B., Crainic, R. & Horodniceanu, F. *C. r. heb. Séance, Acad. Sci., Paris* **III**, 91 (1982).
10. Chow, M. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7518 (1982).
11. Emini, E.A., Jameson, B.A., Lewis, A.J., Larsen, G.R. & Wimmer, E. *J. Virol.* **43**, 997 (1982).
12. Strohmaier, K., Franze, R. & Adam, H.-J. *J. gen. Virol.* **59**, 295 (1982).
13. Bittle, J.L. *et al.* *Nature* **298**, 30 (1982).
14. Pfaff, E., Mussgay, M., Bohm, H.O., Schulz, G.E. & Schaller, H. *EMBO J.* **1**, 869 (1982).
15. Sangar, D.V., Black, D.N., Rowlands, D.J. & Brown, F. *J. gen. Virol.* **35**, 281 (1977).
16. Kupper, H. *et al.* *Nature* **289**, 555 (1981).
17. Boothroyd, J.C. *et al.* *Nature* **290**, 800 (1981).
18. Kleid, D.G. *et al.* *Science* **214**, 1125 (1981).
19. Callis, J.J. & Winther, M.D. Personal communication.
20. Gerin, J.L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 2365 (1983).

tance to neutralization is confined to an eight-amino-acid tract specified by the sequence of bases 279–302 from the start of the region coding for VP1. Furthermore they have been able to identify the precise amino acids required for the neutralization of the virus. These observations imply that the eight-amino-acid sequence on VP1 represents the site on the virus involved in virus neutralization.

This, with the earlier work on foot-and-mouth disease virus, clearly points the way to a new approach to the production of virus vaccines. With those virus diseases for which an effective and cheap vaccine is already available such an approach may not be of commercial importance. However, for the several virus diseases for which no vaccine is available, because of the present inability to grow the virus in

sufficient quantity, the synthetic approach, whether by the organic or biosynthetic route, clearly has attractions, particularly if, as the poliovirus work implies, a peptide consisting of only eight amino acids can stimulate protective levels of neutralizing antibody. The recent promising work on hepatitis B by Lerner's group provides a good example of such an application²⁰. Clearly the key problem is to identify the antigenic site involved in the production of protective antibody. The problem has apparently been solved with foot-and-mouth disease and polio viruses. It will be surprising if the success achieved with these does not stimulate similar work with other viruses. □

F. Brown is at the Wellcome Research Laboratories, Pirbright, Woking, Surrey GU24 0NE.

Evolutionary biology

Extinctions, catastrophic and gradual

from Jared M. Diamond

TODAY, we accept without question that extinctions have occurred. They provide a subject of theoretical interest to ecologists and of practical interest to conservationists. We can hardly imagine what a shock the discovery of bones of extinct dinosaurs, ape-men and mammoths caused for thinkers of the early nineteenth century, accustomed to viewing species as immutable and immortal.

These discoveries quickly led to the development of two opposing theories to explain extinctions of fossil species. The catastrophist theory stems from Baron Cuvier, who surmised the past occurrence of 'terrible events' that decimated species quickly. The gradualist theory instead posited slow effects of progressive climatic changes. More recently, Cuppy reasoned that 'the Age of Reptiles ended because it had gone on long enough and it was all a mistake in the first place' (that is, dinosaurs were disgustingly ugly and deserved extinction)¹. Participants at a recent symposium* ignored Cuppy's theory but reappraised the catastrophist and gradualist views in the light of a broad spectrum of extinction studies².

The Raup–Sepkoski plot³ of the frequency of extinctions against geological time displays a steady background rate, on which are superimposed occasional sharp peaks. Walter Alvarez (University of California, Berkeley) summarized the status of the ultimate catastrophist theory that he proposed in 1980 for one of these peaks. According to his theory, the mass extinctions at the Cretaceous–Tertiary (C–T) boundary 65 Myr ago were due to a collision of a 10-km asteroid with the Earth⁴. Consideration of the number, size distribution and orbits of the Earth-

crossing Apollo asteroids leads one to expect collision of a 10-km object with Earth about every 100 Myr. The geological evidence that this did happen at the C–T boundary is twofold: anomalously very high concentrations of iridium in a worldwide thin layer at the boundary, the ratios of iridium, gold, platinum and osmium corresponding to an object of chondritic composition; and altered microtektites, indicative of an impact, in the same thin layer. This layer exactly marks the sudden extinction of most species of calcareous planktonic foraminifera and coccoliths, many brachiopods and bryozoans, and the last ammonites. The layer approximately but less certainly marks the last dinosaurs. Possible collision-induced mechanisms of these extinctions are at least four: complete darkness and cessation of photosynthesis for months, because of the blocking of sunlight by atmospheric dust; immediate cooling, for the same reason; delayed heating, because of the greenhouse effect of atmospheric water vapour if the impact was in the ocean; and generation of nitric acid by shock heating of the atmosphere. Other iridium anomalies and microtektite layers may mark collisions of other large or smaller objects, producing worldwide or local extinctions at other times.

Like Alvarez, S. Stanley (Johns Hopkins University) and A. Knoll (Harvard University) viewed extinctions in geological time, but they focused respectively on shallow-water marine molluscs and on vascular plants. Mass extinctions of molluscs have been variously attributed to the lowering of sea level with contraction of the continental shelves, or to temperature drops; Stanley argued for the latter view. The pattern of vascular plant extinctions contrasts

with that for animals. Clear evidence for decimation of vascular plants at the C–T boundary is lacking. The great radiations of animal groups immediately followed the great extinctions of other animal groups: for example, the radiation of mammals following the dinosaur extinctions, and numerous other replacements, at the C–T boundary. In contrast, the successive radiations of the trimerophytes, progymnosperms, pteridosperms and angiosperms immediately preceded the extinctions of the rhyniophytes, trimerophytes, progymnosperms and many previously important gymnosperm groups respectively. Thus, animal extinctions permitted animal radiations, whereas plant radiations caused plant extinctions.

Knoll related these plant–animal differences in extinction patterns to basic biological differences between plants and animals. Through devices such as wilting during droughts, resprouting from roots and prolonged quiescence as seeds, plants can survive short-term catastrophes (for example, asteroid collisions) far better than animals. On the other hand, plant species are much more uniform in their ecological role (almost all being photosynthesizers) than are animals, which are divided among many guilds and trophic roles. Animals can escape a new competitor by shifting trophic roles, but plants cannot. Thus plants have often been decimated by evolution of new competitors in the past or by human introduction of competitors from other continents in modern times.

Paul Martin (University of Arizona) reviewed the evidence that man's colonization of North and South America, Australia and oceanic islands had produced catastrophic extinction waves. Victims of these waves included the mammoths and other large mammals, the moas and other large birds, and many small insular species^{5,6}. Remains of fossil large mammals showing evidence of death at the hands of man are rare in the New World and unknown in Australia. Martin has interpreted this fact to suggest a *blitzkrieg*-swift extinction wave by the first arriving hunters, leaving few butchered carcasses as evidence⁷. Gradualists instead consider this fact to argue that the extinctions resulted from late-Pleistocene climate changes. Obviously, it becomes crucial to discover whether the large mammals disappeared quickly on man's arrival or survived for thousands of years.

One is forced to the latter conclusion if one accepts all published carbon-14 dates for the extinct North American mammals. However, the objects analysed differ widely in the care that was taken to eliminate contamination by younger carbon, or in the firmness of association with an extinct mammal's body. If one restricts consideration to the presumed collagen fraction ex-

*The Sixth Annual Systematics Symposium, held on May 14 1983 at the Field Museum of Natural History in Chicago, was organized by Matthew Nitecki of the Museum.

tracted from bone, the occasional late dates reported for whole bone disappear, and dates based on extracted amino acids themselves⁸ cluster around 11,000 yr BP (ref. 9). This corresponds to the date of the first stone-tool horizon documented for the New World, the Clovis horizon. The same date is obtained for the extinction of the Shasta ground sloth based on carbon-14 analyses of its dried dung balls, whose provenance from the animal is unassailable. The last dung balls are dated around 11,000 yr BP, whether taken from caves in juniper-ash woodland at an elevation of 1,800 ft or from caves in spruce-oak woodland at 6,500 ft. If the Shasta ground sloth did succumb to climate changes simultaneously in two such different habitats just at the time that Clovis hunters arrived, this supposedly stupid animal deserves credit for hitherto unsuspected cunning in confusing later palaeontologists.

An extinction symposium held 10 years ago would have considered man as an architect but not a victim of extinctions. The formerly prevalent view was that there has been only a single hominid line, hence no extinctions of hominid species. There is now widespread agreement that two hominid lines co-existed in Africa until almost 1 Myr ago: the *Homo habilis*-*Homo erectus* line, that survived as *Homo sapiens*; and the *Australopithecus robustus* line, that became extinct (A. Walker, Johns Hopkins University). The latter line was a small-brained vegetarian with massive chewing muscles and teeth. (It is unclear whether other smaller hominid skulls represent female *H. erectus* or a third line that also became extinct.)

Walker saw the evolution of *Homo* and the extinction of *A. robustus* as related to changes in three African large-animal guilds between 2 and 1 Myr BP: the scavengers, stalking carnivores and running carnivores. During this period the carnivore guilds suffered the extinctions of the running hyena *Euryboas*, three sabretooth cats and a cheetah; *H. habilis* evolved to join and then leave the brown and spotted hyenas in the scavenger guild; *H. erectus* and the lion evolved to join the leopard in the stalking-carnivore guild; and the hunting dog evolved to join the cheetah in the running-carnivore guild¹⁰. By the time the dust from this evolutionary reshuffle had settled around 1 Myr BP, the three guilds had reached their modern composition, *H. erectus* had become the first hominid to expand beyond Africa (to Asia) and the prey species¹¹ *A. robustus* was gone.

It is possible, but disputed, that another hominid extinction occurred later: that of Neanderthal man. European Neanderthals either evolved very quickly into, or were replaced very quickly by, *H. sapiens*. Walker's analysis of skull shape shows that Neanderthals were much more similar in this respect to *H. erectus* than to *H. sapiens*. Does this suggest another

bifurcation and another extinction?

The constant background of extinctions, from which these occasional extinction catastrophes stand out, was reviewed by J. Diamond (University of California, Los Angeles) and illustrated for Rocky Mountain mammals by B. Patterson (Field Museum). Background extinctions reveal themselves on several time scales: over a decade or two, in the turnover of populations on islands¹²; over a century, in the extinctions of populations in patches of habitats fragmented by man^{13,14}; over the past ten millennia, in the extinctions of populations on land-bridge islands or habitats fragmented by late-Pleistocene changes in sea level and climate^{15,16}; and over geological time, in the extinctions implicit in patterns of endemism¹². Theoretical considerations suggest that the risk of extinction for isolated populations in a fluctuating environment should decrease with population density, island size, generation time, population stability, intrinsic rate of increase and the ratio of birth rates to death rates¹⁷. The latter two predictions remain untested, but the first four are confirmed by the studies at various time scales. The steep decrease in extinction rate with population density means that isolated populations of carnivores, large animals and habitat specialists are generally at greater risk than populations of herbivores, small animals and habitat generalists.

Finally, T. Lovejoy (World Wildlife Fund) described an experimental study that World Wildlife Fund is carrying out jointly with the Brazilian government to examine the effects of clearing Amazonian rain forest for agriculture¹⁸. The results of this study will be described in detail in a later article in *News and Views*. What World Wildlife Fund and the Brazilian government are doing as a small-scale controlled experiment in order to learn, man throughout the world is doing as a massive uncontrolled experiment in the name of development. Today's asteroid collision is of our own making. □

Jared Diamond is Professor of Physiology at University of California Medical School, Los Angeles, California 90024.

1. Cuppy, W. *How to Become Extinct* (Farrar & Rinehart, New York, 1941).
2. *Extinctions* (ed. Nitecki, M.H.) (University of Chicago Press, in the press).
3. Raup, D.M. & Sepkoski, J.J. *Science* **215**, 1501 (1982).
4. Alvarez, W. *et al. Science* **208**, 1095 (1980).
5. Martin, P.S. *Science* **179**, 969 (1973).
6. Diamond, J.M. *Nature* **298**, 787 (1982).
7. Mosimann, J.E. & Martin, P.S. *Am. Sci.* **63**, 304 (1975).
8. Ho, T.Y., Marcus, L.F. & Berger, R. *Science* **164**, 1051 (1969).
9. Thompson, R.S. *et al. Quat. Res.* **14**, 360 (1980).
10. *Evolution of African Mammals* (eds Maglio & Cooke) (Harvard University Press, 1978).
11. Brain, C.K. *The Hunters or the Hunted?* (University of Chicago Press, 1981).
12. Diamond, J.M. *Proc. 19th int. Orn. Congr.* 777 (1980).
13. Whitcomb, R.F. *et al. in Forest Island Dynamics in Man-dominated Landscapes* (eds Burgess & Sharpe) (Springer, New York, 1981).
14. Willis, E.O. *Proc. 19th int. Orn. Congr.* 783 (1980).
15. Patterson, B.D. *Southwestern Nat.* **25**, 33 (1980).
16. Brown, J.H. *Am. Nat.* **105**, 467 (1971).
17. Leigh, E. *J. theor. Biol.* **90**, 213 (1981).
18. Lovejoy, T. E. *Parks* **5**, 13 (1980).

Ecology

Cycling index

from John T. Finn

IN Robert May's *News and Views* article on energy cycling¹, an error has been occasioned by the use of different definitions of the cycling index in the earlier and later papers he discusses. In our early papers^{2,3} we defined the cycling index (as described by May) as 'the fraction of total flow through the system that derives from cycling, expressed as a ratio to the fraction of the total that derives from flow straight through the system'. This index can vary from zero to infinity. However, in my later papers^{4,5}, I redefined the cycling index as the fraction of total flow through the system that derives from cycling, expressed as a ratio to the total flow through the system. This index can vary from zero (no cycling) to one (complete cycling). In comparing nitrogen cycling of the Luquillo tropical rain forest in Puerto Rico and Hubbard Brook, May¹ used figures derived from the first definition for Luquillo (1.76) and figures derived from the second definition for Hubbard Brook (0.76), to conclude erroneously that Luquillo cycles more than Hubbard Brook. Using the second definition for comparison with Hubbard Brook, the proper value for Luquillo is 0.48, so according to these models, Hubbard Brook cycles more nitrogen (proportionately) than does the Luquillo rain forest.

At first, it seems that the models must be wrong. Indeed, some components of the gaseous flux of nitrogen at Hubbard Brook and at Luquillo have yet to be adequately measured, so the cycling indices may change. However, it is still instructive to wonder why a temperate forest should cycle more nitrogen than a tropical forest. First, the large forest floor of the temperate forest gives it an advantage in holding nutrients, especially for a forest like Hubbard Brook that is still accumulating both live and dead biomass. Second, higher temperatures at Luquillo speed up processes of nitrogen loss (decomposition and denitrification) as well as gain (fixation). Finally, greater rainfall subjects Luquillo to a higher potential loss of nutrients through leaching. The Puerto Rican rain forest may cycle nutrients faster by using primarily active biological mechanisms to prevent nitrogen from leaching away, but the proportion of nitrogen cycled compared with nitrogen flowing through the system could be less than at Hubbard Brook. □

John T. Finn is Associate Professor of Systems Ecology at the University of Massachusetts, Amherst, Massachusetts 01003.

1. May, R.M. *Nature* **292**, 105 (1981).
2. Finn, J.T. *J. theor. Biol.* **56**, 363 (1976).
3. Patten, B.C. *et al. in Systems Analysis and Simulation in Ecology* (ed. Patten, B.C.) 457 (Academic, New York, 1976).
4. Finn, J.T. *DOE Symp.* **45**, 138 (National Technical Information Service, Springfield, 1978).
5. Finn, J.T. *Ecology* **61**, 562 (1980).

Palaeontology

Triassic 'scenarios'

from M. Howgate

THE tetrapod land faunas of the early part of the Triassic period were dominated by mammal-like reptiles. Herds of herbivorous dicynodonts were being preyed upon by top carnivores such as *Cynognathus*. However by the end of the period this entire fauna had all but vanished to be replaced in all the large-body terrestrial ecological niches by the early dinosaurs and their relatives — the archosaurs. The problems posed by this major faunal replacement and the disparate assemblage of theories it has thrown up were addressed by Alan Charig, from the British Museum (Natural History), at a recent symposium*.

Charig started with a defence of the 'scenario' as a valid method of interpreting past events — a method dismissed by some as mere evolutionary storytelling. By 'scenario' Charig means "a sequential narrative of hypothesized prehistorical events based on conjectural interpretations of palaeontological data". Such a scenario could be deemed to be scientific provided that it were based on a firm foundation of empirical evidence, the interpretations made being the most parsimonious possible and the scenario as formulated open to subsequent modification.

It was on these latter points that Charig himself opposed some of the existing scenarios constructed to explain the faunal transition, on the grounds that they relied too much on "the conjectured effects of conjectured environmental conditions upon conjectured physiological characteristics". The theories he particularly had in mind were those involving hot-blooded dinosaurs and their ilk, that seek to explain the dominance of the archosaurian fauna in terms of their hypothetical ectothermic constitution; or conversely for cold-blooded dinosaurs, theories that depend on a hypothetical deterioration in climate. Similar objections apply to those theories which produce a non-competitive scenario by relying solely on extrinsic factors, such as changes in climate and flora to eliminate the mammal-like reptiles.

Charig's own position is that the replacement was dominated by a competition for food resources (actively or passively — it makes no difference) in which the primary selective factor, and the only one for which we have direct evidence in the fossil record, was the improvement in locomotion in the archosaurian lineage. Thus while the archosaurs attained first a 'semi-improved' and then a 'fully improved' dinosaurian

gait, the mammal-like therapsids, despite alleged improvements in thermal regulation, never achieved a comparable stance and gait.

Throughout the Early Triassic then, a gradual replacement took place in which archosaurian carnivores out-competed their cynodont adversaries, the latter becoming smaller, fewer and less diverse as the replacement 'snowballed' during Middle Triassic times. By mid-Ischigualasto times (Carnian of Argentina) the terrestrial fauna was one of carnivorous pseudosuchian archosaurs preying on a herbivorous assemblage composed of dicynodonts, those previously carnivorous cynodonts which had succeeded in making the transition to herbivory — the traver-

sodontids — and the aberrant rhynchosaurs [now separated from the rhynchoccephalians and placed among the archosauromorphs by Mike Benton (Nature Conservancy Council, Newbury) in another paper at the same symposium].

The post-Ischigualasto formations reveal a second wave of replacement, this time among the herbivores, and Charig sees his scenario developing as follows. The archosaurs, some now with a fully improved gait and hence dinosaurs, over-exploit the therapsid herbivores and rhynchosaurs and drive them to extinction. Then faced with a drastically reduced food supply but an abundance of vegetation, selection pressure drove some of the omnivorous archosaurs to switch to a vegetable diet. So by the beginning of the Jurassic period a new balance had been attained in which both carnivores and herbivores had the same fully improved gait. □

M. Howgate is at the Department of Zoology, University College, London WC1E 6BT.

Immunology

The importance of T3 in the activation of T lymphocytes

from Peter Beverley

MONOCLONAL antibodies against leukocyte surface antigens were first used for identifying and separating subsets of cells, but more recently attention has been focused on the molecules identified by the antibodies. Clearly, if an antibody modifies a cellular function this implies that the molecule to which the antibody binds may be involved in that function. Monoclonal antibodies thus provide probes for investigation of function at the molecular level.

The anti-T3 monoclonal antibody OKT3 was first described as an antibody recognizing all T cells¹, and it has since been used, along with others of the same specificity, to identify and enumerate human T cells in health and disease. Early biochemical data identified T3 as a 20,000-molecular weight glycoprotein with some heterogeneity of charge², but T3 might have remained for some time a molecule looking for a function had it not been shown that anti-T3 antibody had profound effects on T-cell function.

Anti-T3 antibody in very low concentrations is a powerful mitogen for human T cells³ and at the least this observation identifies T3 as a molecule (or molecules) likely to play a part in the activation of the cells. That binding of a ligand to T3 is not sufficient in itself to induce division in resting T cells is shown by the requirement for accessory cells, demonstrated previously by cell depletion and reconstitution experiments³, and confirmed in this issue of *Nature* by the identification of a genetically

determined defect in the ability of some human non-T cells to provide accessory function for T-cell responses to mouse IgG1 anti-T3 antibodies⁴.

Anti-T3 has since been shown to block a variety of T-cell functions including the proliferative response to soluble antigens and both the generation in mixed leukocyte culture and expression of T-cell cytotoxicity⁵. While the mitogenic effect of the antibody made interpretation of the proliferative data difficult, the blocking of the effector phase of cytotoxicity was not open to this criticism and, taken with the mitogenic effect of anti-T3, strongly implied that the T3 molecule was involved in T-cell recognition or triggering by antigen. Recent experiments reported in *Nature* and elsewhere have amply confirmed this view.

The results can be summarized as follows. If interleukin-2-dependent T-cell clones are exposed to anti-T3 antibody, T3 is rapidly lost from the cell surface by a process known as modulation. Cells in which modulation has occurred are no longer capable of responding by proliferation to alloantigen⁶ or soluble antigen⁷ nor, in the case of cytotoxic clones, can they kill appropriate target cells. That loss of T3 is linked to loss of antigen-specific receptors is implied by the parallel loss of a putative T-cell receptor detected by a clone-specific monoclonal antibody⁷. Similarly, T3 is modulated if cells are exposed to the anti-clonotype antibody or to soluble antigen, again implying that the antigen-specific receptor and T3 are linked. Modulated

*A symposium on 'The Structure, Development and Evolution of Reptiles' was held in honour of Professor Angus Bellairs by the Zoological Society of London on 26–28 May 1983.

T-cell clones can still respond to interleukin-2, which suggests that T3 is involved in initial triggering rather than maintenance of clonal expansion.

While the present results have begun to define a role for T3 in somewhat artificial conditions *in vitro*, its physiological role in triggering resting cells is less clear, as is its exact relationship to the recently defined clonotypic molecules. New biochemical data however show that T3 is a complex of polypeptides of several different molecular weights. While the principal component has a molecular weight of 20,000, other molecular species of molecular weights 25,000–28,000, 37,000 and 44,000 have been reported⁸. Furthermore, at least one polypeptide with a molecular weight of 20,000 seems to be non-glycosylated. How these various forms are related to each other and to the clonotypic structures is not yet clear but peptide mapping suggests that at least one 20,000-molecular weight form and the 25,000–28,000-molecular weight form are distinct polypeptides. That T3 is made up of several distinct polypeptides suggests that there may be more than one T3 gene, a possibility supported by the difficulties encountered in attempts to obtain stable expression of T3 in TK⁺ L cells co-transformed with human DNA and herpes virus TK⁺ gene⁹.

The obvious functional importance of T3 has led to much effort in unravelling its structure, physiology and molecular biology, but one additional major problem remains; so far there have been no published reports of homologues in the mouse or rat, although many of the other human T-cell antigens have rodent counterparts. This may be because T3 is a particularly highly conserved molecule so that raising antisera between different mouse strains or mouse and rat is difficult. The recent human data, however, are likely to provoke a more determined search for murine T3. In any event, even if T3 is lacking in mice, in man it is clearly a molecular complex which plays an important part in activation of human T lymphocytes, and the close association of T3 and the T-cell receptor suggests that at the very least T3 is a hallmark of a functionally mature T lymphocyte. □

Peter Beverley is in the ICRF Tumour Immunology Group, University College School of Medicine, University Street, London WC1E 6JJ.

1. Kung, P.C., Goldstein, G., Reinherz, E.L. & Schlossman, S.F. *Science* **206**, 347 (1979).
2. van Anghoven, A., Terhorst, C., Reinherz, E.L. & Schlossman, S.F. *Eur. J. Immun.* **11**, 18 (1981).
3. van Wauwe, J.P.V., De Mey, J.R. & Goossens, J.G. *J. Immun.* **124**, 2708 (1980).
4. Tax, W.J.M., Willems, H.W., Reekers, P.P.M., Capel, P.J.A. & Koene, R.A.P. *Nature* **303**, 445 (1983).
5. Reinherz, E.L., Hussey, R.E. & Schlossman, S.F. *Eur. J. Immun.* **10**, 758 (1980).
6. Meuer, S.C. et al. *J. exp. Med.* **157**, 705 (1983).
7. Zanders, E.D., Lamb, J.R., Feldmann, M., Green, N. & Beverley, P.C.L. *Nature* **303**, 625 (1983).
8. Borst, J., Alexander, S., Elder, J. & Terhorst, C. *J. biol. Chem.* **258**, 5135 (1983).
9. Kavathas, P. & Herzenberg, L.A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 524 (1983).

Astronomy

Asteroid Thisbe's diameter

from David W. Hughes

OCCULTATIONS yield the most accurate estimations of the diameters of asteroids and the technique is well illustrated by considering the twelve observations taken on the night of 7 October 1981 when the asteroid 88 Thisbe passed in front of a ninth-magnitude star in Sagittarius known as SAO187124.

The prediction of this event was made by Gordon Taylor of the Royal Greenwich Observatory, Herstmonceux. Using simply the catalogue position of the star and the published ephemeris of Thisbe, the ground track of the event (the region of the Earth from which the occultation would be seen) was found to stretch from west to east across western Canada. Photographic plates taken on 27 September and 4 October 1981 with the 50-cm Carnegie Double Astrograph on Mount Hamilton and on 6 October 1981 with the 155-cm Astrometric Reflector at the Flagstaff Station, USNO, indicated a more southerly track. This was about 300 km wide and stretched from California through Nevada, Utah, Colorado, Wyoming, South Dakota, Nebraska, Minnesota, Iowa and on to Ontario.

Thisbe's 'shadow' moved along the track from west to east. The occultation was expected to last for a maximum of 9.7 s. As Thisbe had a visual magnitude of 11.8 as opposed to SAO187124's magnitude of 9, the change in stellar brightness caused by the occultation was large and easily detectable by visual observers.

Twelve groups set out to make the observation using telescopes ranging from the 60-cm reflector at Sommers-Bausch Observatory to relatively small portable instruments. The results of their endeavours have recently been reported in *The Astronomical Journal* (**88**, 229; 1983). The authors are R. Millis, L. Wasserman, O. Franz, N. White and E. Bowell (Lowell Observatory, Flagstaff), A. Klemola (Lick Observatory), R. Elliott, W. Smethells and P. Price (Casey Observatory), C. McKay and D. Steel (Sommers-Bausch Observatory) and E. Everhart and E. M. Everhart (Chamberlain Observatory).

Ideally the observers should be placed at equal intervals across the path. A footnote to the paper is worth quoting because it illustrates beautifully the difficulties in observing occultations.

"The referee of this paper asked that we explain our strategy for placing two portable telescopes on essentially the same chord. While we might argue that this was done to conclusively establish the reality of any satellites that might be detected, such was not the case. Final predictions, based on USNO plates, were not available for this occultation until nine hours before the event. In the headlong rush to reach the track from Flagstaff, two of our teams lost contact with each other. As a result, they were deployed in accordance with Murphy's Law."

In essence each observer should obtain

the time of onset and end of the occultation. Each defines a point on the limb of the asteroid and together the observers give a chord across its Earth-pointing face. Visual observers report times which are systematically late relative to times obtained by photoelectric means but this can be corrected for in the analysis.

The first step is to try and fit a circular limb profile to the chord. This was found to have a diameter of 221.8 ± 1.4 km. The true limb profile might depart markedly from a circle and unfortunately only limited coverage of the southern hemisphere of Thisbe was obtained during the 7 October occultation. A least-squares fit of an ellipse to the data points gave a semiminor axis of 110.8 ± 1.1 km and a semimajor axis of $115.3 (+3.5, -9.0)$ km. More north-south coverage is needed to distinguish meaningfully between a circular or elliptical solution, so the authors chose the solution involving the fewest free parameters — the circle — and suggest an uncertainty of ± 4 per cent.

Only the face pointing towards Earth at the time of occultation has had its diameter measured. It is possible that the cross-section of the asteroid varies as a function of rotation and aspect angle. Most well observed asteroids show little or no variation in polarization as a function of rotational aspect and this is taken to indicate that the surface texture and albedo is constant over the body. Thisbe has been assumed to obey this general rule. But the brightness of Thisbe varies by 0.19 magnitudes with a period of 6.0422 h. The 7 October occultation occurred near the time of minimum brightness and therefore minimum cross-sectional area. A 0.19-magnitude variation in brightness is equivalent to an increase in the effective diameter by about 9 per cent. The mean diameter of Thisbe is thus 232 km and the conservative uncertainty is ± 12 km. D. Morrison and B. Zellner (in *Asteroids*, ed. T. Gehrels, p.1090, University of Arizona Press; 1979) estimated the diameter as 210 km by comparing the visual and infrared radiation coming from the asteroid.

Some observations of asteroid occultations have indicated the occurrence of secondary events, which have been attributed to satellites of the primary asteroid blocking out the star light. T. Van Flandern, E. Tedesco and R. Binzel (in *Asteroids*, p.443) have gone so far as to suggest that "satellites are both numerous and commonplace". No secondary occultations were reported during the Thisbe event on 7 October 1981. □

David W. Hughes is a Lecturer in Physics and Astronomy at the University of Sheffield, Sheffield S3 7RH.

Evolution of oncogenes

From *c-abl* to *v-abl*

The discovery that the oncogenes of retroviruses are derived from cellular genes has led to a search for differences between the viral gene and its cellular counterpart that might explain the oncogenic activity of the former. Two oncogenes, the *abl* and the *src* genes, have recently come under such scrutiny. The *src* gene will be the subject of a later article: here, Jean Yin Jen Wang discusses the *abl* gene.

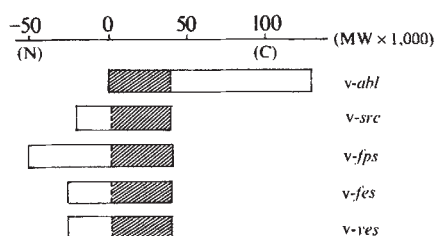
ABELSON murine leukaemia virus (A-MuLV) is an acute oncogenic retrovirus which induces lymphosarcoma in mice and transforms fibroblasts and pre-B lymphocytes in culture. The genome of A-MuLV contains a sequence derived from the mouse genome: by convention, this host-derived sequence is called *v-abl* and the corresponding cellular gene (proto-oncogene) is called *c-abl*. The genome of A-MuLV is a hybrid of *v-abl* and Moloney leukaemia virus (MLV) in which the *v-abl*-coding sequence is linked to a portion of the MLV genome and the expression of this coding sequence is controlled by the promoter-enhancer of MLV. An obvious approach to the study of how *v-abl* induces oncogenic transformation is to compare it with the *c-abl* gene whose product is a normal constituent of the cell.

MLV has a strong promoter: viral RNA produced from an integrated viral genome constitutes about 1 per cent of the total poly(A)-containing RNA of the cells in transformed fibroblasts or pre-B lymphocytes¹. At the same time, cells transformed by A-MuLV express *c-abl* RNA from the normal cellular gene at levels 15 times lower than those of the *v-abl* product; *c-abl* expression does not seem to be affected by the production of the viral RNA¹. It has not yet been determined however whether the increased expression of the viral sequence contributes to the oncogenicity of *v-abl*.

The mouse *c-abl* gene is split into at least 10 exons which are present in *v-abl* as a correctly spliced unit². Thus, *v-abl* can be viewed as a cDNA copy of the *c-abl* RNA. A cDNA clone containing 1.5 kilobases (kb) of the 3' end of a *c-abl* RNA has been obtained which shows that *c-abl* RNA extends beyond the 3' end of *v-abl* by 850 base pairs (bp). The difference between *v-abl* and *c-abl* RNA has been analysed by primer extension experiments whose results indicate that approximately 350 bp on the 5' end of the *c-abl* RNA is also absent from the A-MuLV genome¹. Therefore, *v-abl* is an internal fragment of the *c-abl* RNA. The nucleotide sequence of the *c-abl* DNA shows that *v-abl* begins within an exon and some coding sequences on the 5' end of the *c-abl* gene are deleted from the A-MuLV genome². There is moreover at least one base mutation in *v-abl* which results in an amino acid substitution. Thus *v-abl* differs from *c-abl* both by deletions and by point mutations.

The hybrid genome of A-MuLV gives rise to a fusion protein in which the first 30,000 molecular weight is derived from the MLV *gag* gene and the remaining

130,000 molecular weight from *v-abl*. Expression of the *v-abl*-coding sequence alone in *Escherichia coli* has shown that it encodes a tyrosine-specific protein kinase which phosphorylates itself and bacterial proteins³. The A-MuLV-transformed cells have been shown to contain high levels of phosphotyrosine which can therefore be attributed to the tyrosine kinase activity of the *v-abl*-encoded protein. Not all the *v-abl*-coding sequence is necessary to encode this kinase — expression of different segments of the *v-abl*-coding sequence in bacteria shows that 1.2 kb on the 5' end of *v-abl* is sufficient⁴.



Amino acid homology of five tyrosine kinases.

The nucleotide sequence of *v-abl* has been determined⁵. Although *v-abl* and *v-src* do not cross-hybridize, the amino acid sequence derived from the 1.2-kb kinase-coding region of *v-abl* is about 50 per cent homologous to the C-terminal region of pp60^{src}, which is also a tyrosine kinase. Three other tyrosine kinases encoded by *v-yes*, *v-fps* and *v-fes* share this amino acid homology as well⁶⁻⁸. A schematic diagram of the homologous region of these five tyrosine kinases is shown in the figure. Because the coding sequence for this homologous region is sufficient for the production of an active kinase, this conserved protein sequence can be viewed as a tyrosine kinase domain which is connected to other different sequences in each of the five proteins.

The MLV *gag* sequence is not necessary for the tyrosine kinase activity of the A-MuLV fusion protein and 27,000 out of the 30,000-molecular weight *gag* protein can be eliminated without any effect on the ability of A-MuLV to transform fibroblast⁹. However, while the 160,000-molecular weight fusion protein can transform pre-B lymphocytes present in bone marrow, the *gag*-130,000-molecular weight protein fails to induce bone marrow transformation in culture. To demonstrate the necessity of *gag* in the transformation of lymphocytes, R. Prywes constructed another set of A-MuLV variants which either contain 30,000-molecular weight (*gag*⁺) or

3,000-molecular weight (*gag*⁻) of *gag* sequence linked to the first 2.0 kb of *v-abl*. Clones of both variant DNA can transform NIH 3T3 cells with equal efficiency and viral stocks containing these two variant genomes have comparable focus-forming titres. When these viral stocks were given to bone marrow cells in culture, the *gag*⁺ virus produced transformed colonies but the *gag*⁻ one did not. These results show that while 90 per cent of the MLV *gag* sequence in A-MuLV is not necessary for the transformation of 3T3 cells, 10 per cent of it is not enough for lymphoid transformation. Strictly speaking, therefore, the oncogene of A-MuLV for pre-B lymphocytes is the *gag/v-abl* hybrid rather than *v-abl* itself — whereas *v-abl* alone is probably sufficient for the transformation of NIH 3T3 fibroblasts.

In fact, not all of *v-abl* is required for oncogenicity. As described above, the *gag*⁺ variant containing only the 5' 2.0 kb of *v-abl* can transform both fibroblasts and bone marrow cells. The smallest A-MuLV with transforming ability contains 100 bp of *gag* (molecular weight 3,000) and only the 5' 1.2 kb, that is the kinase-coding region, of *v-abl*⁹. It seems that all it takes is the *v-abl* tyrosine kinase to transform 3T3 cells.

Because the *v-abl*-coding sequence is derived from *c-abl*, it is reasonable to assume that *c-abl*-encoded protein is also a tyrosine kinase with another polypeptide sequence on its C-terminal end. Although the function of that C-terminal region of the *c-abl* protein is unknown, it must be physiologically significant. However, that function is totally dispensable for the transforming activity of *v-abl*.

Since the normal physiological function of *c-abl* is not involved in the oncogenicity of *v-abl*, the tyrosine kinase activity may cause transformation by two different mechanisms. The *v-abl* kinase may have lost its specificity due to its mutations or its higher expression so that it phosphorylates protein(s) which should not be so modified and the mistake(s) triggers transformation. Or, the *v-abl* kinase phosphorylates its physiological substrate(s) (those of the *c-abl* kinase) in an unregulated fashion due to its mutations and this results in transformation. Although we still do not understand the mechanism of transformation, the comparison between *v-abl* and *c-abl* has clearly indicated the direction to be taken towards this end: to study the tyrosine-specific protein kinase. □

Jean Yin Jen Wang is at the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139.

1. Wang, J.Y.J. & Baltimore, D. *Molec. cell. Biol.* (in the press).
2. Wang, J.Y.J. *et al.* (in preparation).
3. Wang, J.Y.J., Queen, C. & Baltimore, D. *J. biol. Chem.* **257**, 13181 (1982).
4. Wang, J.Y.J. & Baltimore, D. (in preparation).
5. Lee, R. *et al.* (in preparation).
6. Kitamura, N. *et al.* *Nature* **297**, 205 (1982).
7. Shibuya, M. & Hanafusa, H. *Cell* **30**, 787 (1982).
8. Hempe, A. *et al.* *Cell* **30**, 775 (1982).
9. Prywes, R., Foulkes, J.G., Rosenberg, N. & Baltimore, D. (submitted).

Magnetic flux tubes on the Sun

H. C. Spruit* & B. Roberts

Department of Applied Mathematics, University of St Andrews, St Andrews, Fife KY16 9SS, UK

Magnetic fields are the cause of almost all forms of solar activity. Near the solar surface, and possibly in the entire convection zone, these fields occur in the form of isolated flux tubes. In recent years, new views have been developed (and older ones revived) in which this property plays a central role. Here we review these ideas, dealing with the nature of the solar cycle, sunspot structure, the origin of spicules and the source of mechanical heating in the solar atmosphere. The ideas are illustrated with the aid of a simple mathematical model for the behaviour of thin magnetic flux tubes. The properties of inhomogeneities in the corona (coronal loops) are also discussed.

THE Sun is permeated by magnetic field. Its presence is responsible, directly or indirectly, for much of what is seen in the solar atmosphere. The magnetic field introduces structure into the atmosphere in which it resides, transforming what might otherwise be a relatively amorphous plasma into one that exhibits many spectacular phenomena, such as the sunspot, the flare and the prominence. By analogy, it is evident that many stars will possess similar or even more exotic structures than those seen on the Sun; the Sun provides us with an invaluable testing ground for our understanding of a cosmic plasma.

There are several different ways in which a magnetic field influences the solar plasma. In the convection zone the field inhibits heat transport, resulting in cool sunspots. In the visible layers (photosphere) of the Sun the field may act as a wave guide, transporting mechanical energy from the convection zone to the atmosphere above. In the higher layers, the energy stored in the field is released, heating the chromosphere and corona. Occasionally the energy stored temporarily in the field is released rapidly, resulting in a solar flare.

An important aspect of solar activity is the extreme inhomogeneity of the magnetic field and/or the plasma contained in it. In the photosphere, this inhomogeneity manifests itself in the form of isolated flux tubes, embedded in a nearly field-free fluid. In the inner corona the field is space filling and magnetic forces dictate the behaviour of the plasma in it to a large extent. The visible structures (coronal loops, spicules) are the result of strong density and temperature inhomogeneities, aligned with, and confined by, the field. Still farther away from the solar surface, in the outer corona, the solar wind dominates, carrying plasma and magnetic field radially away from the Sun.

Intense flux tubes

Apart from sunspots, intense flux tubes comprise almost all of the magnetic flux seen on the solar surface¹⁻³. These tubes have typical dimensions that are near (and possibly below) the presently available resolution limit (about 300 km) and possess strong fields (1–2 kG). Higher in the atmosphere, they rapidly expand to merge with their neighbours (see Figs 1 and 2). It is commonly believed that all stars with convective outer envelopes have magnetic fields with a similar inhomogeneous distribution. Theoretically, this inhomogeneity is interpreted as a consequence of the interaction between convection and the magnetic field⁴⁻⁷. This interaction, as it takes place in an incompressible unstratified fluid, has been studied in great detail using numerical as well as analytical methods⁷. These calculations have clearly demonstrated the formation of flux tubes in a highly conducting magnetized fluid. Schematically, one can say that turbulence and magnetic fields in a fluid of high electrical conductivity expel each other. The Lorentz forces suppress

turbulence within the field, and (three-dimensional) turbulence expels fields from the centres of turbulent elements towards their boundaries. Recently, the process has also been studied numerically in conditions closely approximating those near the solar surface⁸.

The theoretical study of solar magnetic fields is made difficult by their inhomogeneity, and by the strong stratification of the plasma due to gravity. As a result many aspects of solar activity can only be studied using numerical methods. It has been found, however, that several interesting physical questions can also be answered using a simple approximate model for the structure of magnetic flux tubes. In this model, the so-called thin tube approximation⁹⁻¹⁸, the field is conceived to exist in the form of slender bundles of field lines (flux tubes) embedded in a field-free fluid. Mechanical equilibrium between the tube and its surroundings is ensured by a reduction of the gas pressure inside the tube, which compensates for the force exerted by the magnetic field. Diffusion of the magnetic field due to ohmic resistance is neglected in this model. The motion of the gas inside the tube, and of the tube as a whole through the external gas, can be calculated relatively easily, and the complications of compressibility of the gas, gravity, and the stratification of the atmosphere can be taken into account. The thin tube approximation is useful up to the level in the atmosphere where the tubes have fanned out and merged with each other. Above this level the continuous nature of the field must be taken into account. Because the thin tube model is simple and easily applicable to realistic solar conditions it is taken as the main point of reference for this review. We note, however, that this biases the discussion and urge the reader to consult the references for aspects of solar activity not mentioned here.

We consider first a thin vertical magnetic flux tube embedded in a field-free environment. The wave modes of such a tube

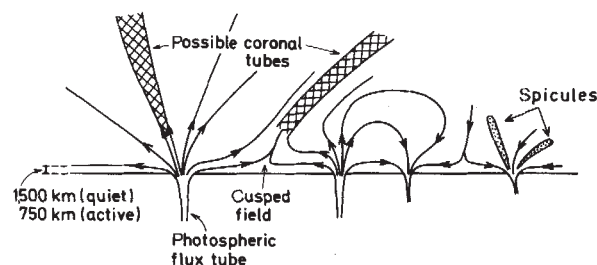


Fig. 1 The small scale magnetic field near the solar surface. At the surface, observations (such as those of Fig. 2) show that the field is concentrated in thin tubes with a field strength of $\sim 1,500$ G. At this level the gas between the tubes appears to be field-free, though a possible weak field of ≤ 5 G cannot be excluded at present. The fields of the individual tubes expand in the atmosphere above the surface, and merge with each other to form the coronal magnetic field. The chromosphere, with its spicules, is located around this merging level.

* Permanent address: Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D-8046 Garching bei München, FRG.

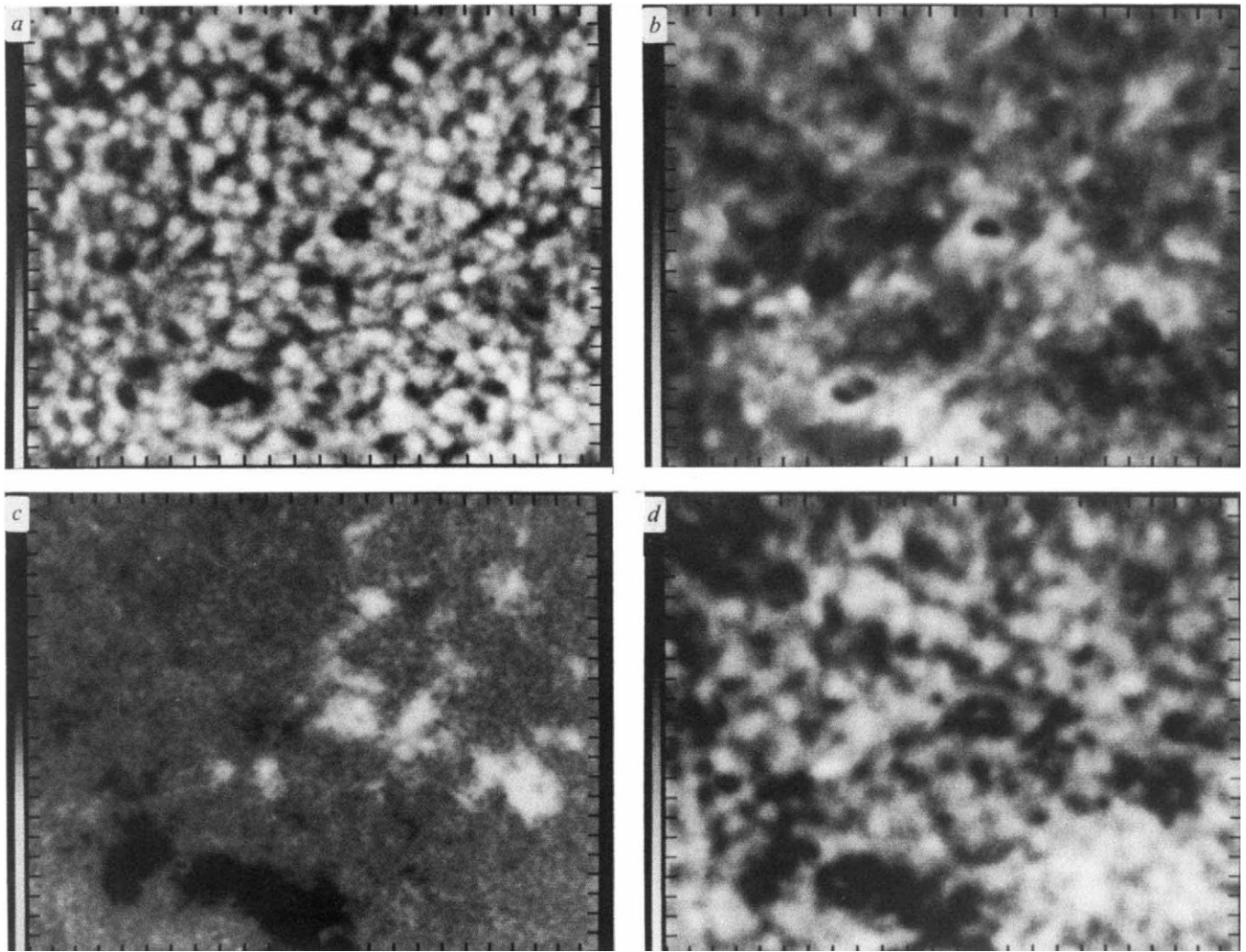


Fig. 2 High-resolution observations of a young active region near the centre of the solar disk, showing: *a*, the continuum intensity (near 525 nm); *b*, the intensity in the Fe I line at 525.022 nm; *c*, the magnetic field strength as measured in this line (white is positive, black negative polarity); and *d*, the line of sight component of the velocity (white directed towards, dark away from the observer). The dark patch in the lower left hand corner is a sunspot. Tick marks are 2 arc s (1,500 km on the Sun) apart. The magnetic fields are located preferentially in the intergranular lanes (dark in *a*) and have a downward flow associated with them. By extrapolation of the field above and below the surface (using theoretical as well as observational methods), a schematic picture as shown in Fig. 1 results. These observations, which show magnetic structures at the highest resolution obtained thus far (~ 0.4 arc s) were obtained by Drs A. M. Title and T. D. Tarbell at Sacramento Peak Observatory, with the Lockheed tunable filter. (Courtesy of T. D. Tarbell, Lockheed Palo Alto Research Laboratory.)

are most readily understood if we first neglect gravity. In a magnetic tube with mass density ρ , field strength B , gas pressure p , sound speed $c = (\gamma p / \rho)^{1/2}$, Alfvén speed $a = B / (\mu \rho)^{1/2}$, embedded in a fluid of density ρ_e , there are three waves. They are analogous to the modes of oscillation of an elastic string.^{11,16-18} One mode is a torsional wave propagating at the Alfvén speed, another is the longitudinal wave propagating with the speed $c_1 = ca(c^2 + a^2)^{-1/2}$, and the third is a transversal (kink) wave propagating with speed $c_k = a[\rho / (\rho + \rho_e)]^{1/2}$. The longitudinal wave consists of alternating constrictions and expansions, and the transversal wave of kinks of the tube as a whole.

When gravity is not neglected, the cross-sectional area of the tube increases with height. This, together with the appearance of a buoyancy force, changes the character of the longitudinal and the transversal tube waves, which become dispersive. The torsional wave is unaffected. For the simple case of an isothermal atmosphere with pressure scale-height H , both the longitudinal and transversal tube waves depend on height (z) and time (t) as $\exp(i\omega t - ikz + z/4H)$. The frequency ω and the wavenumber k are related by

$$\omega^2 = c_p^2 k^2 + \Omega^2 \quad (1)$$

where c_p is the mode's propagation speed (c_k or c_1) and Ω its cutoff frequency¹¹⁻¹³

$$\Omega_1 = \frac{c_1}{4H} \left(9 - \frac{8}{\gamma} + \frac{16(\gamma-1)}{\gamma^2} \frac{c^2}{a^2} \right)^{1/2}; \quad \Omega_k = \frac{c_k}{4H} \quad (2)$$

We have assumed here that the propagation is adiabatic. The dispersive nature of the waves is analogous to that of an acoustic wave propagating vertically in a field-free atmosphere. For such a wave, $c_p = c$ and $\Omega = \Omega_a = c/2H$. Its amplitude increases with height as $\exp(z/2H)$, twice as fast as the tube waves. The difference is due to the increase of the tube's cross-section with height.

The above dispersion relation shows that frequencies above the cutoff Ω are propagating ($k^2 > 0$) and low frequencies are evanescent ($k^2 < 0$). An evanescent wave can be regarded as the quasistatic response of the flux tube to slow changes in the conditions at its base. In addition to looking at a disturbance of fixed frequency ω , we can consider a disturbance due to an impulsive source¹⁹. The resulting disturbance then consists of a wave front moving with the speed c_p trailed by a 'wake', oscillating at the frequency Ω .

Fig. 3 Spicules at the solar limb; these are jets of cool plasma protruding into the hot corona (invisible on this picture). The image is 60,000 km wide. (Courtesy of R. B. Dunn, Association of Universities for Research in Astronomy Inc., Sacramento Peak Observatory.)



It has been realized since the late 1950s that there is a close relation between heating of the chromosphere and corona, and the presence of magnetic fields^{20,21}. Apparently, the field provides a channel by which mechanical energy generated in the convection zone is transported to the atmosphere. It is now generally accepted that most of the atmospheric heating occurs through such a process (except in layers close to the photosphere), although its precise nature is less clear. One candidate is the continual reconnection of field lines in the atmosphere, as a consequence of the rearrangement of the footpoints of the field at the photosphere^{22,23}. Another possibility is the dissipation of magnetic waves induced in the field by motions in the convection zone. Probably both processes contribute. Which kind of wave (if any) is most likely to be important depends on the efficiency with which it is generated, the ease with which it propagates and the way in which it dissipates. We now consider these questions from the flux tube point of view. The properties of waves in a homogeneous field in a stratified atmosphere have been discussed elsewhere^{24,25}.

Waves in flux tubes on the Sun are most likely to be generated by the action of the surrounding granulation which buffets and jostles the tubes. The generation will not be equally effective for the three modes, due to differences in their propagation characteristics¹³. Pressure fluctuations excite longitudinal tube waves, horizontal displacements of the tube in the granular velocity field cause transversal waves. It is uncertain if torsional Alfvén waves can be excited by granulation, but they may be generated by the unwinding of a twisted tube as it is buoyed up through the convection zone during the eruption of an active region^{22,26}.

For the propagation of waves, the existence of a cutoff frequency is of crucial importance. Motions will be excited in the convection zone over a broad range in frequencies, but only those frequencies above the cutoff frequency of the wave give rise to significant amplitudes in the higher layers. In the solar atmosphere the cutoff frequency Ω_c for acoustic waves is $\sim 3 \times 10^{-2} \text{ rad s}^{-1}$. In a typical flux tube, where $c \approx a$, the cutoff for the longitudinal and transversal wave is $\Omega_l \approx \Omega_a$ and $\Omega_k \approx \Omega_a/3$, respectively. The frequencies occurring in the granular velocities are around $1.3 \times 10^{-2} \text{ rad s}^{-1}$. This is just above the cutoff of the transversal wave, but well below the cutoff of acoustic waves and longitudinal tube waves. Thus we expect to see transversal waves at high amplitudes in the chromosphere¹³, and indeed strong transversal motions are observed in H_α fibrils²⁷⁻²⁹. The longitudinal tube waves originating in the convection zone are likely to show up at much lower amplitudes than the transversal waves, and this too seems in accord with observations³⁰. Radiative damping, known to be important for acoustic waves, will similarly occur in longitudinal tube waves³¹ (although it is less strong than in the acoustic case). By contrast, the transversal waves and the torsional waves are unaffected by such damping. The general question of the dissipation of such waves has hardly been explored yet.

Up to this point we have considered only the properties of the tube modes in an isothermal atmosphere. When the tube is embedded in a thermally unstable stratification, such as the solar convection zone, the longitudinal mode may become unstable. The instability is driven by the buoyancy force in much the same way as in ordinary convection. The magnetic

field exerts a stabilizing influence, and constrains the unstable motion to a flow along the tube³²⁻³⁶. If the instability sets in as a downdraft, the tube is partially evacuated, leading to a compression of the field. The instability thus leads to a new equilibrium state of higher field strength, but with a lower total energy. The magnetic energy is increased at the expense of the gravitational energy of the gas within the tube. A calculation³⁵ of this equilibrium in conditions appropriate for the Sun gives field strengths at the $\tau = 1$ level of about 1.5 kG, consistent with observed values^{37,38}.

Spicules

A phenomenon that is probably connected with the dynamics of flux tubes is the spicule. Spicules are jet-like structures in which gas moves upward into the atmosphere with high speeds ($20\text{--}30 \text{ km s}^{-1}$) and with almost uniform density³⁹. They have been observed for over a century and are the most characteristic feature of the chromosphere (see Fig. 3). Their motion is quite remarkable, resembling a steady flow rather than a ballistic motion or a shock wave.

Recent calculations⁴⁰⁻⁴³ support the idea that spicules are a highly nonlinear manifestation of wave motions in a flux tube. A wave in a vertical flux tube, if it is generated impulsively enough in the photosphere, will quickly grow in amplitude as it propagates (adiabatically) up the tube. In linear theory, the disturbance consists of a wavefront trailing an oscillating wake¹⁹. When the amplitude of the wave approaches the speed of sound, strong longitudinal motions are generated nonlinearly, even if such motions were absent at the source of the disturbance. As in linear theory, these longitudinal flows have the form of a wavefront trailing an oscillating wake, both now strongly nonlinear. Numerical studies in which a torsional (Alfvén) wave⁴¹ or a longitudinal tube wave⁴³ is launched at the photosphere show that the generated motions strongly resemble spicules, provided that the source is a sufficiently strong pulse of sufficiently short duration. Although it has not been tested yet, it seems likely that a transversal flux tube wave generated at the photosphere will produce similar effects. As this wave (as discussed earlier) is generated more readily and amplifies more easily with height than the torsional or the longitudinal waves, the transversal wave may be a likely candidate for the disturbance which ultimately produces spicules. However, even if this turns out to be the case, some serious questions about spicules still remain to be resolved, most notably their remarkable *in situ* disappearance, and the general question of their radiation and ionization physics.

Sunspots

Sunspots (see Fig. 4) are cool, dark patches on the solar surface; they are regions of strong magnetic field. With their large sizes ($3 \times 10^3\text{--}6 \times 10^4 \text{ km}$), they represent the opposite extreme to the subtelescopic flux tubes that make up the small scale magnetic field on the Sun. Field strengths in spots are typically 3,000 G, significantly larger than the strengths of small flux tubes. Spots have been observed for at least 2,000 yr, but their magnetic nature was only discovered in 1908⁴⁴. Recently, the possibility that sunspots may have been absent for long periods of recorded history, most notably during the Maunder minimum (1645–1715), has attracted much attention⁴⁵.

The structure of the magnetic field of a spot in the layers below the visible surface is uncertain. The uniform appearance of a sunspot umbra has naturally led to the suggestion, implicit in most discussions of sunspot structure, that a spot extends into deeper layers as a single coherent tube. For several reasons, this picture is becoming less attractive. It has been known for some time⁴⁶ that the formation of a spot occurs through the accumulation of a large number of small tubes. Similarly, the slow decay of a spot by the 'evaporation' of small tubes from its boundary is a well-established process⁴⁷. These observations suggest that a spot might actually be a large collection of tiny flux tubes. What appears to be a uniform structure at the surface may look more like a bundle of threads deeper down^{48,49}. Such a model has been proposed by Parker⁵⁰.

Theoretical support for this idea has come through an analysis by Meyer *et al.*⁵¹, who studied the stability of a spot against splitting into smaller tubes. It was found that a spot, in the form of a single vertical flux tube, is stable to such splitting in the upper layers of the convection zone (down to 5,000 km depth). The stability in these layers is related to the buoyancy of magnetic fields. Due to the lower gas density inside the field, a flux tube which is held fixed at some depth in the convection zone will assume a vertical position. Similarly, a cluster of tubes held fixed at some depth will stay together by buoyancy in spite of the repelling forces between the field lines. An analogy for such a cluster would be a bunch of tethered balloons⁵¹. The balloons would be pressed together, representing the homogeneous field of the observed spot. The stability analysis also showed, however, that a single tube would be unstable to splitting in deeper layers. Here, the field would thus be in the form of separate tubes (representing the strings of the balloons). The depth at which the tubes first separate could be about 1,000 km, that is, slightly below the visible layers of the spot. This could explain the presence of umbral dots and umbral granulation (small scale inhomogeneities in temperature) seen in most spots.

It has been pointed out⁵² that, if a spot of the type described is to be in mechanical equilibrium, an as yet unidentified force must exist which at some depth keeps the tube together (the hand holding the strings). Hydrodynamical effects have been proposed for this force^{50,51}. It is, however, not necessary to assume that a spot is everywhere in mechanical equilibrium during its life. The time scale for establishing a magnetic equilibrium is L/a , where L and a are a characteristic length and Alfvén speed. Near the solar surface this time scale is short (15 min) compared with the life time of the spot (days to months), but in deeper layers this is not the case. If spots extend down to the base of the convection zone, and if the field strength there is of the order 10^4 G, it takes one month for a disturbance in the field to propagate from this depth to the surface. The reason is that at large depths the density of the gas is high and consequently its response to the magnetic forces much slower than near the surface. Thus, the observed evolution of spots may just reflect the absence of a magnetic equilibrium state at their roots. It is important in this picture that these roots be located near the base of the convection zone. This location would agree with presently popular (but still rather qualitative) ideas which favour the interface between the convection zone and the radiative interior as the location where the dynamo process takes place, and where the toroidal field is stored from which active regions emerge⁵⁴⁻⁵⁶.

Accurate measurements of the solar luminosity⁵⁷ have recently shown variations which are just as large as expected when sunspots were dark patches without any compensating brighter areas elsewhere on the surface. The heat blocked by spots is apparently stored elsewhere. Calculations of the thermal response of a convection zone to spots on its surface^{58,59} have shown that this storage and the absence of bright rings are natural consequences of the thermal properties of a convective stellar envelope (see also ref. 60). The storage time scale is large, of the order of 10^5 yr.

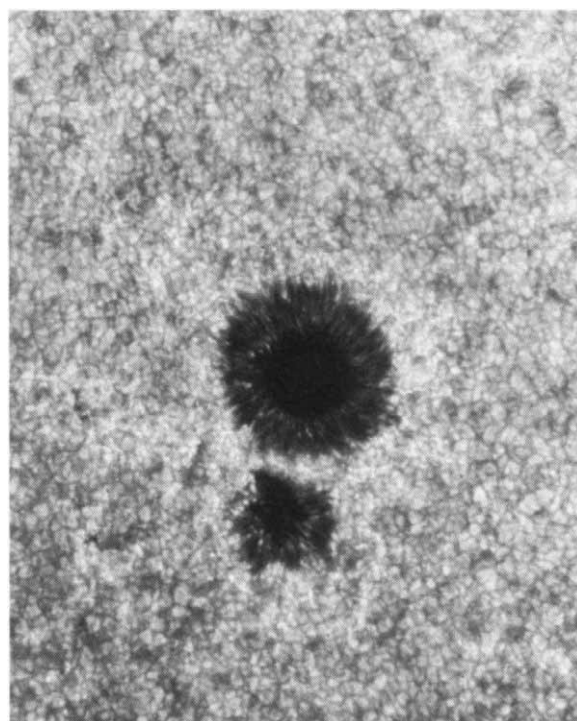


Fig. 4 Sunspots. Note the penumbral fine structure and the surrounding granulation. (Courtesy of R. Muller, Pic du Midi Observatory.)

Activity cycle

Explanations of the 22-yr activity cycle of the Sun have generally invoked a dynamo process^{61,62}, by which magnetic flux is generated in the convection zone. In mathematically tractable dynamo models several complicating factors must be left out. Usually ignored are: the flux tube nature of the magnetic field, which may also persist within the convection zone; the very strong buoyancy of horizontal magnetic fields; and the back-reaction of magnetic forces on the convective flow. Investigations that include aspects of the discrete nature of the field have been made^{63,64}. In numerical simulations, the backreaction of the field on the flow has been included⁶⁵. It has been realized⁶⁶ that strong buoyancy forces act in magnetic flux tubes of the required field strength. Under their influence tubes either float to the surface as a whole or they form loops of which the tops break through the surface and the lower parts descend to the bottom of the convection zone. The convection zone, being unstable itself, enhances this process^{66,67}. Small tubes take longer to erupt through the surface because they feel stronger drag forces. Detailed calculations of the process⁶⁸ show that even a tube with the size of a very small spot, if located within the convection zone, will erupt in less than a year. Yet, to produce an activity cycle, horizontal fields are needed below the surface which survive for about 10 yr. A possible solution is that these fields are located at the boundary between the convection zone and the radiative interior. Since the interior is very stably stratified, all fields floating down under the influence of convective buoyancy forces accumulate at this boundary. The idea that the field is concentrated near the bottom of the convection zone has been advanced several times in the past^{54,55,66,69,70}, for various reasons including the above. Another reason is the effect (topological pumping) that in a convecting layer magnetic fields tend to be pushed downward^{71,72}. If the field is indeed stored at the boundary, the solar cycle would be maintained by an 'interface dynamo', as opposed to the conventional view in which dynamo action occurs dispersed throughout the convection zone. A qualitative picture of the magnetic field in such a model is given in Fig. 5. An interesting consequence of this model is that in a fully convective star (lacking the stabilizing radiative core) the activity cycle, if

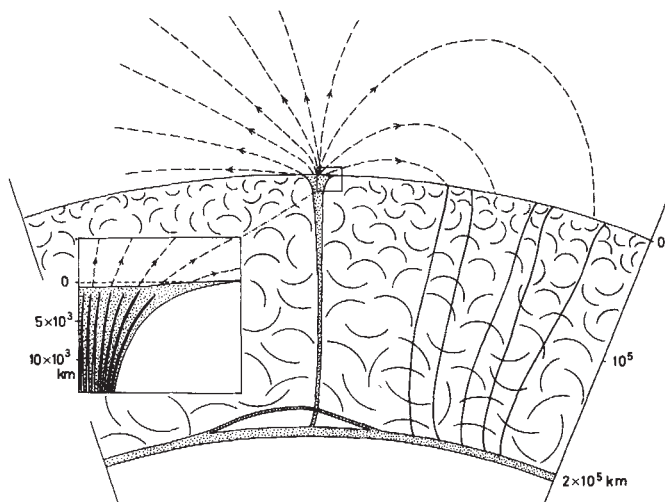


Fig. 5 Sketch of the magnetic field in the convection zone of the Sun, under the 'interface dynamo' assumption. Horizontal fields are stored at the base of the convection zone during the cycle. Active regions form from sections brought up by buoyancy (one shown in the process of rising). After eruption through the solar surface a nearly potential field is set up in the atmosphere (broken lines), connecting to the base of the convection zone via almost vertical flux tubes. Hypothetical small scale structure of a sunspot is shown in the inset.

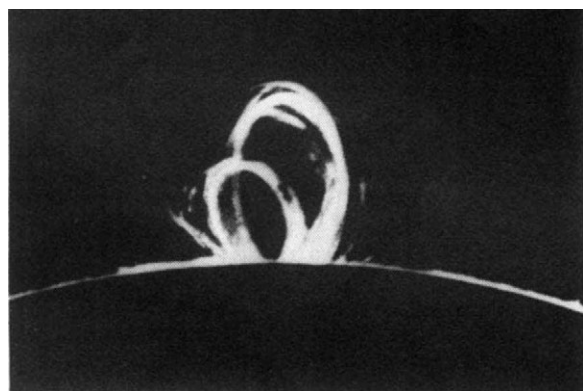


Fig. 6 Postflare loops in the corona, tracing out magnetic field lines. (Courtesy of Association of Universities for Research in Astronomy Inc., Sacramento Peak Observatory.)

present, would be qualitatively different. A tantalizing observation in this connection is that the magnetic activity of main sequence stars with spectral type later than M5 is apparently strongly reduced⁷³. These stars are just those that are expected to be fully convective.

Coronal loops

The magnetic structuring so evident in the photosphere is also apparent in the corona^{74,75}, although there it takes on a quite different guise. The magnetic field pervades all of the corona and dominates over the gas pressure, so coronal phenomena (such as flares and prominences) are primarily the result of magnetic processes. Above active regions the coronal magnetic field is fairly strong (say, 100 G) and the temperature is high ($2-3 \times 10^6$ K). High-resolution pictures of the corona in active regions show it to be made up of many individual tube-like structures known as coronal loops. They are made visible not because of large variations in magnetic strength—the field is roughly uniform—but through density and temperature inhomogeneities which are aligned with the field. The coronal magnetic field strongly inhibits the conduction of heat across the field-lines, so a loop is thermally isolated from its neighbours. Conduction of heat along a field-line, however, is very effective. Thus, local input of heat at some point on a field-line results in the heating of plasma along the entire field-line. The

heating leads to 'evaporation' of chromospheric gas at the foot-points of the loop, the plasma density in the loop increases, and with it the radiation from the loop (mostly in the UV). Coronal loops therefore outline field-lines, and indicate localized sources of heat. This is especially evident in a solar flare. In a flare a large amount of energy stored in the magnetic field is released suddenly through an instability which results in field-line reconnection. Some flares disrupt large regions of the corona, others apparently are contained within a single magnetic loop. Recent discussions of flares may be found in refs 75–78. By the evaporation process described above, the flare produces many coronal loops observable in X rays. When the energy source fades away, they cool down and can be seen in visible light as 'post-flare' loops (Fig. 6).

Magnetic field-lines (whether visible as coronal loops or not) are elastic structures and so tend to respond rapidly to local pressure fluctuations. The magnetoacoustic oscillations that may ensue have recently been worked out in some detail^{17,79} and show the presence of two distinct scales of oscillations, one acoustic in nature and the other alfvénic. The fast (alfvénic) oscillations are particularly interesting for several reasons. First, they are closely analogous to Love waves in seismology and Pekeris waves in oceanography; second, their short periods (several to tens of seconds), in typical coronal conditions, suggest that they may be linked to the long observed type IV radio oscillations⁸⁰. These fast oscillations arise as free modes of vibration only in dense loops (the ones that are observed in X rays). They provide a potentially useful diagnostic tool for determining coronal magnetic field strengths, which are otherwise difficult to measure.

Densities and temperatures along a coronal loop are determined by conduction, radiation and mechanical heating. The processes of conduction and radiation in the corona are reasonably well understood, but the nature of the mechanical heating is still uncertain. We should note, however, that there are uncertainties in modelling the thermal structure of loops in the transition region, where temperatures and densities change rapidly from coronal to chromospheric values. These difficulties, of loop modelling and with choice of boundary conditions at the transition region base of loops, remain to be explored fully. (For a discussion of conduction in the transition region, see ref. 81.)

While the precise nature of the mechanical heat input into the corona is not known, it is almost certainly magnetic in form. Sound waves^{82,83} or internal gravity waves⁸⁴ may well contribute to the heating of the low chromosphere, but provide insufficient energy flux to heat the corona^{85,86}. Several mechanisms for magnetic heating of the corona are actively being pursued, including surface magnetohydrodynamic waves^{87,88}, enhanced dissipation of sheared Alfvén waves⁸⁹, resonant processes in coronal loops^{53,90–92}, dislocation of magnetic field⁹³, and anomalous joule heating⁹⁴. No consensus has yet emerged⁹⁵; we must await further developments in theory and observation before the important question as to the nature of coronal heating can be answered with any confidence.

Independently of the precise nature of the heating mechanism, scaling laws have been developed which relate the temperature, density and length of a coronal loop⁹⁶. These may prove useful in analysing observations of magnetic activity in the stars⁹⁷, but observational uncertainties (even for the case of the Sun) suggest the need for caution^{98,99}. The stability of coronal loops, both thermally^{100–102} and dynamically¹⁰³, has been considered extensively in recent years but is still a matter of some debate.

Conclusions

The small-scale nature of the photospheric magnetic field has led to new theoretical interpretations of several classical phenomena such as the sunspot, the spicule and the solar activity cycle. The models for sunspots and for spicules currently under development seem particularly promising. Developments are, however, hampered by the fact that most of the magnetic

structure on the solar surface occurs on scales too small to be resolved with present instruments. Spectacular improvements are expected with the advent of the Solar Optical Telescope, a 2-m space-borne instrument to be launched by NASA towards the end of this decade.

Studies of small scale structures (loops) in the corona are also promising and offer some support for the view that the physics of coronal heating via magnetic fields may be clarified in the not too distant future.

We thank the SERC for financial support.

1. Stenflo, J. O. *IAU Symp.* **71**, 69–99 (1977).
2. Harvey, J. W. in *Highlights of Astronomy* Vol. 4 (ed. Müller, E.) Part 2, 223–239 (Reidel, Dordrecht, 1977).
3. Tarbell, T. D., Title, A. M. & Schoolman, S. A. *Astrophys. J.* **229**, 387–392 (1979).
4. Kraichnan, R. N. *J. Fluid Mech.* **77**, 753–768 (1976).
5. Galloway, D. J., Proctor, M. R. E. & Weiss, N. O. *Nature* **266**, 686–689 (1977).
6. Weiss, N. O. in *Highlights of Astronomy* Vol. 4 (ed. Müller, E.) Part 2, 241–250 (Reidel, Dordrecht, 1977).
7. Proctor, M. R. E. & Weiss, N. O. *Rep. Prog. Phys.* **45**, 1317–1379 (1982).
8. Nordlund, A. in *Solar and Stellar Magnetic Fields: Origins and Coronal Effects* (ed. Stenflo, J. O.) (Reidel, Dordrecht, in the press).
9. Parker, E. N. *Astrophys. J.* **189**, 563–568 (1974); **190**, 429–436 (1974).
10. Parker, E. N. *Cosmical Magnetic Fields* ch. 8–10 (Clarendon, Oxford, 1979).
11. Defouw, R. J. *Astrophys. J.* **209**, 266–269 (1976).
12. Roberts, B. & Webb, A. R. *Solar Phys.* **56**, 5–35 (1978); **64**, 77–92 (1979).
13. Spruit, H. C. *Astr. Astrophys.* **98**, 155–160 (1981).
14. Spruit, H. C. in *Solar Phenomena in Stars and Stellar Systems* (eds Bonnet, R. M. & Dupree, A. K.) 289–300 (Reidel, Dordrecht, 1981).
15. Wilson, P. R. *Astrophys. J.* **221**, 672–676 (1978).
16. Roberts, B. *Solar Phys.* **69**, 27–38 (1981); **69**, 39–56 (1981).
17. Edwin, P. M. & Roberts, B. *Solar Phys.* **76**, 239–259 (1982).
18. Spruit, H. C. *Solar Phys.* **75**, 3–17 (1982).
19. Rae, I. C. & Roberts, B. *Astrophys. J.* **256**, 761–767 (1982).
20. Babcock, H. W. & Babcock, H. D. in *Electromagnetic Phenomena in Cosmical Physics* (ed. Lehnert, B.) 239–247 (Cambridge University Press, 1958).
21. Osterbrock, D. E. *Astrophys. J.* **134**, 347–388 (1961).
22. Parker, E. N. *Astrophys. J.* **191**, 245–254 (1974).
23. Rosner, R., Golub, L., Coppi, B. & Vaiana, G. S. *Astrophys. J.* **222**, 317–332 (1978).
24. Thomas, J. H. A. *Rev. Fluid Mech.* **15**, 321–343 (1983).
25. Schwartz, S. J. & Leroy, B. *Astr. Astrophys.* **112**, 93–103 (1982).
26. Piddington, J. H. *Astrophys. J.* **233**, 727–735 (1979).
27. Giovanelli, R. G. *Solar Phys.* **44**, 299–314 (1975).
28. Bray, R. J. & Loughhead, R. E. *The Solar Chromosphere*, 220 (Chapman & Hall, London, 1974).
29. Wentzel, D. G. *Astr. Astrophys.* **76**, 20–23 (1979).
30. Giovanelli, R. G., Livingston, W. C. & Harvey, J. W. *Solar Phys.* **59**, 49–64 (1978).
31. Roberts, B. & Webb, A. R. *Solar Phys.* **68**, 71–85 (1980); **68**, 87–102 (1980).
32. Parker, E. N. *Astrophys. J.* **221**, 368–377 (1978).
33. Roberts, B. & Webb, A. R. *Solar Phys.* **59**, 249–274 (1978).
34. Spruit, H. C. & Zweibel, E. G. *Solar Phys.* **62**, 15–22 (1979).
35. Spruit, H. C. *Solar Phys.* **61**, 363–378 (1979).
36. Unno, W. & Ando, H. *Geophys. Astrophys. Fluid Dyn.* **12**, 107–115 (1979).
37. Tarbell, T. D. & Title, A. M. *Solar Phys.* **52**, 13–25 (1977).
38. Stenflo, J. O. *Solar Phys.* **32**, 41–63 (1973).
39. Beckers, J. M. A. *Rev. Astr. Astrophys.* **10**, 73–100 (1972).
40. Shibata, K. in *Proc. Japan–France Seminar on Solar Physics* (eds Moriyama, F. & Henoux, J. C.) 50–65 (CNRS, Paris, 1980).
41. Hollweg, J. V., Jackson, S. & Galloway, D. J. *Solar Phys.* **75**, 35–61 (1982).
42. Suematsu, Y., Shibata, K., Nishikawa, T. & Kitai, R. *Solar Phys.* **75**, 99–118 (1982).
43. Hollweg, J. V. *Astrophys. J.* **257**, 345–353 (1982).
44. Hale, G. E. *Astrophys. J.* **28**, 100–101, 315–343 (1908).
45. Eddy, J. et al. *The Solar Output and Its Variation* (ed. White, O. R.) (Colorado Associated University Press, Boulder, 1977).
46. Vrabec, D. in *Chromospheric Fine Structure* (ed. Athay, R. G.) 201–231 (Reidel, Dordrecht, 1974).
47. Harvey, K. & Harvey, J. *Solar Phys.* **28**, 61–71 (1973).
48. Severny, A. B. *Soviet Astr.-AJ* **9**, 171–182 (1965).
49. Piddington, J. H. *Astrophys. Space Sci.* **75**, 273–287 (1981).
50. Parker, E. N. *Astrophys. J.* **230**, 905–913 (1979).
51. Meyer, F., Schmidt, H. U. & Weiss, N. O. *Mon. Not. R. astr. Soc.* **179**, 741–761 (1977).
52. Parker, E. N. *Solar Phys.* **40**, 291–301 (1975).
53. Žugžda, Y. D. & Locāns, V. *Solar Phys.* **76**, 77–108 (1982).
54. Spiegel, E. A. & Weiss, N. O. *Nature* **287**, 616–617 (1980).
55. Galloway, D. J. & Weiss, N. O. *Astrophys. J.* **243**, 945–953 (1981).
56. Spruit, H. C. & Van Ballegoijen, A. A. *Astr. Astrophys.* **106**, 58–66; **113**, 350 (1982).
57. Willson, R. C., Culkis, S., Janssen, M., Hudson, H. S. & Chapman, G. A. *Science* **211**, 700–702 (1981).
58. Spruit, H. C. *Astr. Astrophys.* **108**, 348–355, 356–360 (1982).
59. Foukal, P., Livshits, M. & Fowler, L. *Astrophys. J.* (in the press).
60. Sweet, P. A. *Vistas Astr.* **1**, 679–685 (1955).
61. Cowling, T. G. A. *Rev. Astr. Astrophys.* **19**, 115–135 (1981).
62. Schüßler, M. in *Solar and Stellar Magnetic Fields: Origins and Coronal Effects* (ed. Stenflo, J. O.) (Reidel, Dordrecht, 1983).
63. Schüßler, M. *Nature* **288**, 150–152 (1980).
64. Parker, E. N. *Astrophys. J.* **256**, 302–315 (1980).
65. Gilman, P. A. & Miller, J. *Astrophys. J. Suppl.* **46**, 211–238 (1981).
66. Parker, E. N. *Astrophys. J.* **198**, 205–209 (1975).
67. Golub, L., Rosner, R., Vaiana, G. S. & Weiss, N. O. *Astrophys. J.* **243**, 309–316 (1981).
68. Moreno-Inertis, F. *Astr. Astrophys.* (submitted).
69. Rosner, R. & Vaiana, G. S. in *X-ray Astronomy* (eds Giacconi, R. & Setti, G.) 129–151 (Reidel, Dordrecht, 1980).
70. Van Ballegoijen, A. A. *Astr. Astrophys.* **113**, 99–112 (1982).
71. Drobyshevskii, E. M. & Yuferev, V. S. *J. Fluid Mech.* **65**, 33–44 (1974).
72. Arter, W. thesis, Univ. Cambridge (1983).
73. Golub, L., Giampapa, M. S., & Soderblom, D. in *Activity in Red Dwarf Stars* (ed. Rodono, M.) (Reidel, Dordrecht, in the press).
74. Vaiana, G. S. & Rosner, R. A. *Rev. Astr. Astrophys.* **16**, 393–428 (1978).
75. Priest, E. R. *Solar Magnetohydrodynamics* (Reidel, Dordrecht, 1982).
76. Priest, E. R. (ed.) *Solar Flare Magnetohydrodynamics* (Gordon & Breach, New York, 1981).
77. Švestka, Z. *Solar Flares* (Reidel, Dordrecht, 1976).
78. Spicer, D. & Brown, J. C. in *The Sun as a Star* (ed. Jordan, S.) 413–470 (NASA SP-450, 1981).
79. Edwin, P. M. & Roberts, B. *Solar Phys.* (in the press).
80. Roberts, B., Edwin, P. M., & Benz, A. O. *Nature* (submitted).
81. Shoub, E. C. *Astrophys. J.* **266**, 339–369 (1983).
82. Ulmschneider, P. & Bohn, H. U. *Astr. Astrophys.* **99**, 173–176 (1981).
83. Ulmschneider, P. & Stein, R. F. *Astr. Astrophys.* **106**, 9–13 (1982).
84. Mihalas, B. W. & Toomre, J. *Astrophys. J.* **249**, 349–371 (1981).
85. Athay, R. G. & White, O. R. *Astrophys. J.* **226**, 1135–1139 (1978).
86. Mein, P., Mein, N. & Schmieder, B. in *Proc. Japan–France Seminar on Solar Physics* (eds Moriyama, F. & Henoux, J. C.) 70–76 (CNRS, Paris, 1980).
87. Ionson, J. A. *Astrophys. J.* **226**, 650–673 (1978).
88. Rae, I. C. & Roberts, B. *Geophys. Astrophys. Fluid Dyn.* **18**, 197–226 (1981).
89. Heyvaerts, J. & Priest, E. R. *Astr. Astrophys.* **117**, 220–234 (1983).
90. Hollweg, J. V. *Solar Phys.* **70**, 25–66 (1981).
91. Ionson, J. A. *Astrophys. J.* **254**, 318–334 (1982).
92. Hollweg, J. V. *Astrophys. J.* (in the press).
93. Parker, E. N. *Astrophys. J.* **244**, 631–643 (1981).
94. Rosner, R., Golub, L., Coppi, B. & Vaiana, G. S. *Astrophys. J.* **222**, 317–332 (1978).
95. Kuperus, M., Ionson, J. A. & Spicer, D. S. *Ann. Rev. Astr. Astrophys.* **19**, 7–40 (1981).
96. Rosner, R., Tucker, W. H. & Vaiana, G. S. *Astrophys. J.* **220**, 643–665 (1978).
97. Stenflo, J. O. (ed.) *Solar and Stellar Magnetic Fields: Origin and Coronal Effects* (Reidel, Dordrecht, 1983).
98. Roberts, B. & Frankenthal, S. *Solar Phys.* **68**, 103–109 (1980).
99. Jordan, C. *Astr. Astrophys.* **86**, 355–363 (1980).
100. Antiochos, S. K. *Astrophys. J. Lett.* **232**, L125–L129 (1979).
101. Hood, A. H. & Priest, E. R. *Astr. Astrophys.* **87**, 126–131 (1980).
102. Craig, I. J. D. & McClymont, A. N. *Nature* **294**, 333–334 (1981).
103. Van Hoven, G. in *Solar Flare Magnetohydrodynamics* (ed. Priest, E. R.) 217–275 (Gordon & Breach, New York, 1981).

ARTICLES

Bathymetry estimates in the southern oceans from Seasat altimetry

T. H. Dixon & M. E. Parke

Jet Propulsion Laboratory, Mail Code 183–701, 4800 Oak Grove Drive, Pasadena, California 91109, USA

A mean sea surface height based on Seasat altimeter data has been high-pass filtered to reveal a pattern of geoid anomalies which are highly correlated with sea-floor topography. The technique is used to locate new bathymetric features in the poorly surveyed southern oceans.

MANY parts of the world's oceans have been poorly surveyed, especially the south-west Pacific and south Indian Oceans. New bathymetric information in these areas can be obtained by

studying the mean ocean surface, which is mainly a gravitational equipotential surface (the geoid). The shape of the geoid is correlated with sea-floor topography, and can be described

using radar altimeter data taken from a satellite platform.

The basic altimeter measurement is the distance between the spacecraft and the ocean surface, as inferred from the two-way travel time of a reflected radar pulse. Because of the finite temporal pulse width (3.1 ns), the radar pulse reflects from a surface region (footprint) whose diameter will depend on sea state, varying between 2 and 12 km. The shape and amplitude of the reflected pulse waveform also depends on sea state and in fact can be used to infer wave height and wind speed¹. Determination of the exact travel time requires modeling of the interaction of the radar signal with the ocean surface^{2,3}. The major factors affecting travel time include the radial position of the satellite, atmospheric time delay, the geoid height at the sub-satellite point, and deviations in ocean surface height due to tides, ocean current topography and response to atmospheric pressure variations⁴. Tracking data allow determination of satellite position, and corrections for atmospheric time delay can be applied based on measurement and/or modelling of atmospheric parameters. The mean sea surface height is then by definition the sum of all uncorrected, time-invariant contributions to the altimeter measurement.

The major influence on mean ocean surface elevation is the marine geoid. On basin scales (>5,000 km) the difference between the geoid and the Earth's reference ellipsoid can be as much as 200 m, while over shorter length scales (<500 km) geoid variation ranges up to 30 m. In contrast, variations in ocean current topography are <1 m. With the exception of warm- and cold-core rings that 'spin off' the major western boundary currents (with height signals of up to 1 m) the height contribution of mesoscale eddies is typically <20 cm.

Ocean floor topography has a major effect on the shape of the marine geoid, because of its proximity to the ocean surface and the large density contrast between rock and water. The correlation between the sea surface (geoid) and sea-floor topography is strongest for features having wavelengths (spatial scales) from 30 to 300 km (refs 5–8). Where age constraints for topographic features and underlying crust are available, high-quality altimeter data can be used to predict sea-floor topography to better than ± 500 m along individual altimeter ground tracks⁹.

The Seasat altimeter collected 70 days of data over the 100-day period from 5 July to 10 October 1978. The internal precision of the height measurement was better than 1 part in 10^7 or about 5 cm. However, the position of the satellite was known only to about 1.6 m (r.m.s.). Orbital variation is caused by atmospheric drag, solar radiation pressure and poorly modelled gravitational effects¹⁰, with the latter effect dominating¹¹. Errors in satellite position appear as long-wavelength deviations in the data. That part of the error which changes with time can be removed or minimized with subsequent processing.

Of the total 70-day data set, only about 45 days represent independent ground tracks. Consequently the Seasat data set is not sufficient to construct a good two-dimensional description of the geoid for some of the shorter wavelengths (<100 km) where the geoid and sea-floor topography are known to be correlated. For example, the effect of some seamounts will not be included in the measured mean sea surface, due to sparse track coverage. Nevertheless, parts of the southern oceans are so poorly surveyed that Seasat altimetry provides important new bathymetric constraints. Some new features of seafloor topography and/or geoid anomalies in the south-west Pacific and Indian oceans that have become apparent from the Seasat data are now discussed.

Data

Parke and Stavert¹² have described a method for combining all the Seasat altimeter data into a two-dimensional map of the 'mean sea surface' that is not biased by long wavelength temporal variations, mainly orbit error. Temporal variations are estimated by minimizing altimeter-measured height differences at ground track cross-over points. Since the satellite repeats

the same ground location at these points the geoid height must be the same. The estimated variation is removed and the data are interpolated to a grid. Averaging of data collected over weeks and months will tend to reduce the remaining temporal variations, which may be related to transient oceanographic effects. The resulting surface is a close approximation to the marine geoid, and has a horizontal resolution of ~ 50 km and an internal precision over these scales better than 20 cm r.m.s.

The mean sea surface contains geoid information at all wavelengths. Longer wavelengths (<600 km) include information about deep-seated mass anomalies within the Earth, while shorter wavelengths only reflect shallow structure and sea-floor topography. The magnitude of longer wavelength effects is large enough to obscure shorter wavelength geoid features. In order to investigate the shallow anomalies, longer wavelength trends must be removed from the data.

There are three approaches to removing long wavelength trends in altimeter data. A common technique is based on calculation of derivatives to produce a gradient map^{12,13}. These maps enhance detection of high gradient features such as geoid anomalies caused by seamounts. However, the ability to detect longer wavelength features with more subtle gradients is reduced.

A second approach is to remove a reference surface from the data. Good reference surfaces for this application are recent Goddard Earth Models (GEMs) based on surface gravity measurements, such as GEM-10B¹⁴. However, GEM-10B does not extend beyond degree and order 36. Removing this surface would still leave wavelengths up to 2,000 km in the data, significantly longer than those of interest here. In addition, geoid models are not perfect, especially in the southern oceans where surface data are sparse or questionable¹⁵. This leads to metre-level errors in precisely those areas we seek to investigate.

The third approach is to use a high-pass filter, such as removal of a running mean. Such a filter is an internally consistent trend removal process. As it is independent of other data sets it is not biased by variation in data quality or coverage. High-pass filters do have the unfortunate habit of introducing side lobes adjacent to high-amplitude anomalies. However, GEMs ideally represent a filtered version of the true geoid and thus can also introduce side lobes in detrended data.

For this study, the Seasat mean sea surface¹² was high-pass filtered by subtracting an average height value for a $6^\circ \times 6^\circ$ square region centred on each point. This band pass enables investigation of all residual geoid anomalies which may be correlated with seafloor topography, that is those with wavelengths <600 km. The resulting surface was contoured between ± 4 m. The contour interval clips high amplitudes to emphasize subtle mid-range features. Sub-regions in the South Pacific and Indian oceans are shown in detail in Fig. 1. A compilation of existing bathymetry¹⁶ for the same sub-regions is shown in Fig. 2.

The high-pass filter used for this study minimizes edge effects near land and preserves subtle features in geoid structure at continental margins, smaller boundary seas, and even in lakes. The side lobes introduced by the filter are of the order of 15% of the amplitude of the main geoid anomaly and have the opposite sense. An extreme example occurs in the vicinity of the Tonga-Kermadec trench, where the geoid anomaly ranges down to -20 m, and associated side lobes range to $+3$ m within 100 km of the trench. Side lobes are not a general problem, since the unknown or poorly surveyed areas discussed here typically have local geoid anomalies less than ± 1.5 m. Associated side lobes have a maximum amplitude (0.2 m) similar to data precision, and much smaller than the contour interval.

The geoid anomaly map in Fig. 1 represents the first opportunity to investigate synoptically major bathymetric features in the southern oceans. We caution, however, that bathymetry inferred from geoid anomalies alone may be incorrect, inasmuch as density anomalies in the crust and upper mantle may exist with little or no topographic expression (see, for example, ref.

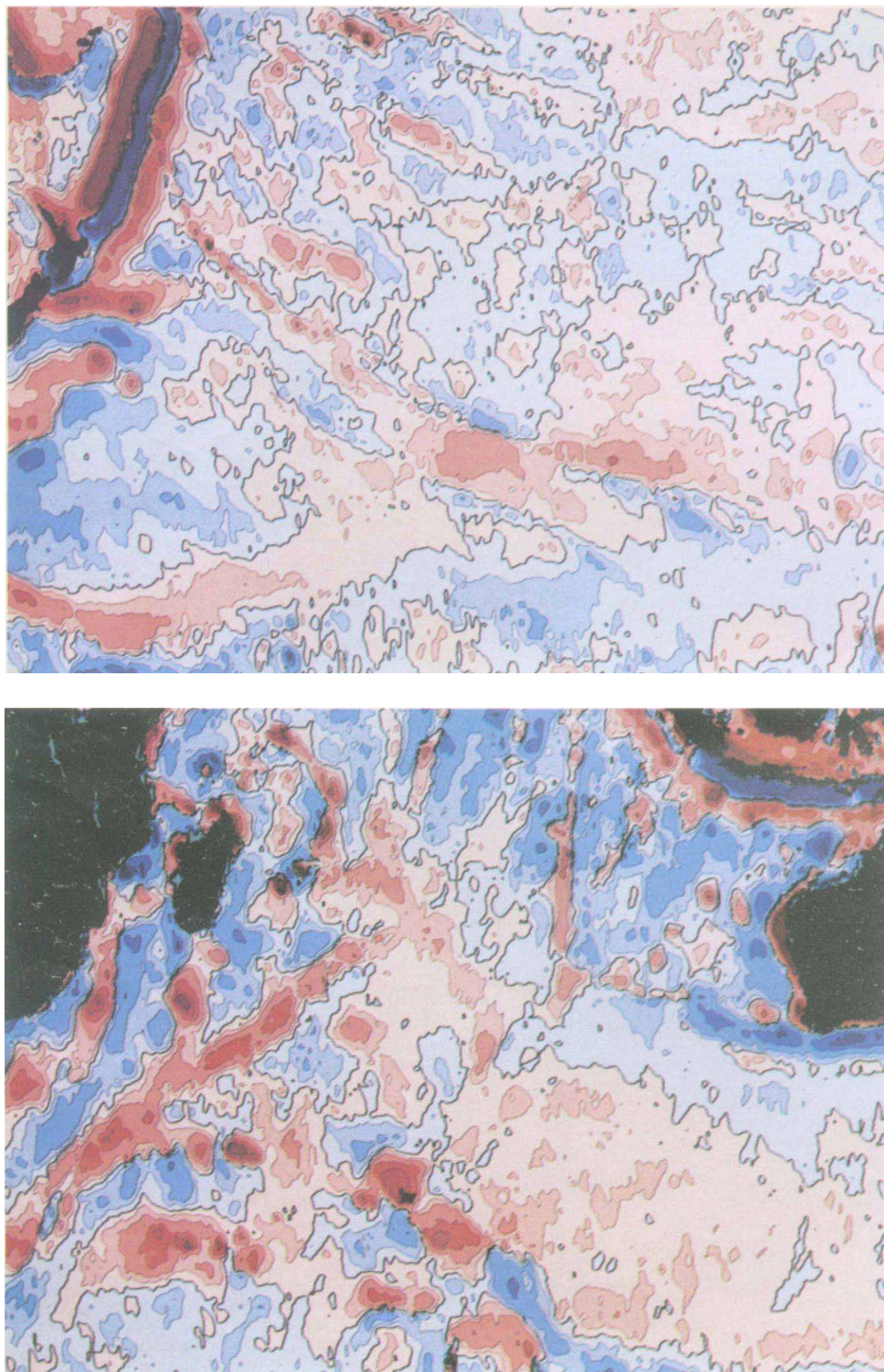


Fig. 1 Geoid anomaly map derived from a Seasat mean sea surface¹², high-pass filtered to remove long (>600 km) wavelengths. The anomalies are contoured at 1-m increments from -4 m to +4 m, except for the middle range -1 m to +1 m, which is contoured at 50-cm increments. Darkest blue represents areas deeper than -4 m, brightest red indicates areas higher than +4 m, and intermediate colour changes occur at each contour interval. The zero contour, between lightest blue and pale pink or cream, is indicated with a heavy black line. *a*, South Pacific Ocean; *b*, South Indian Ocean.

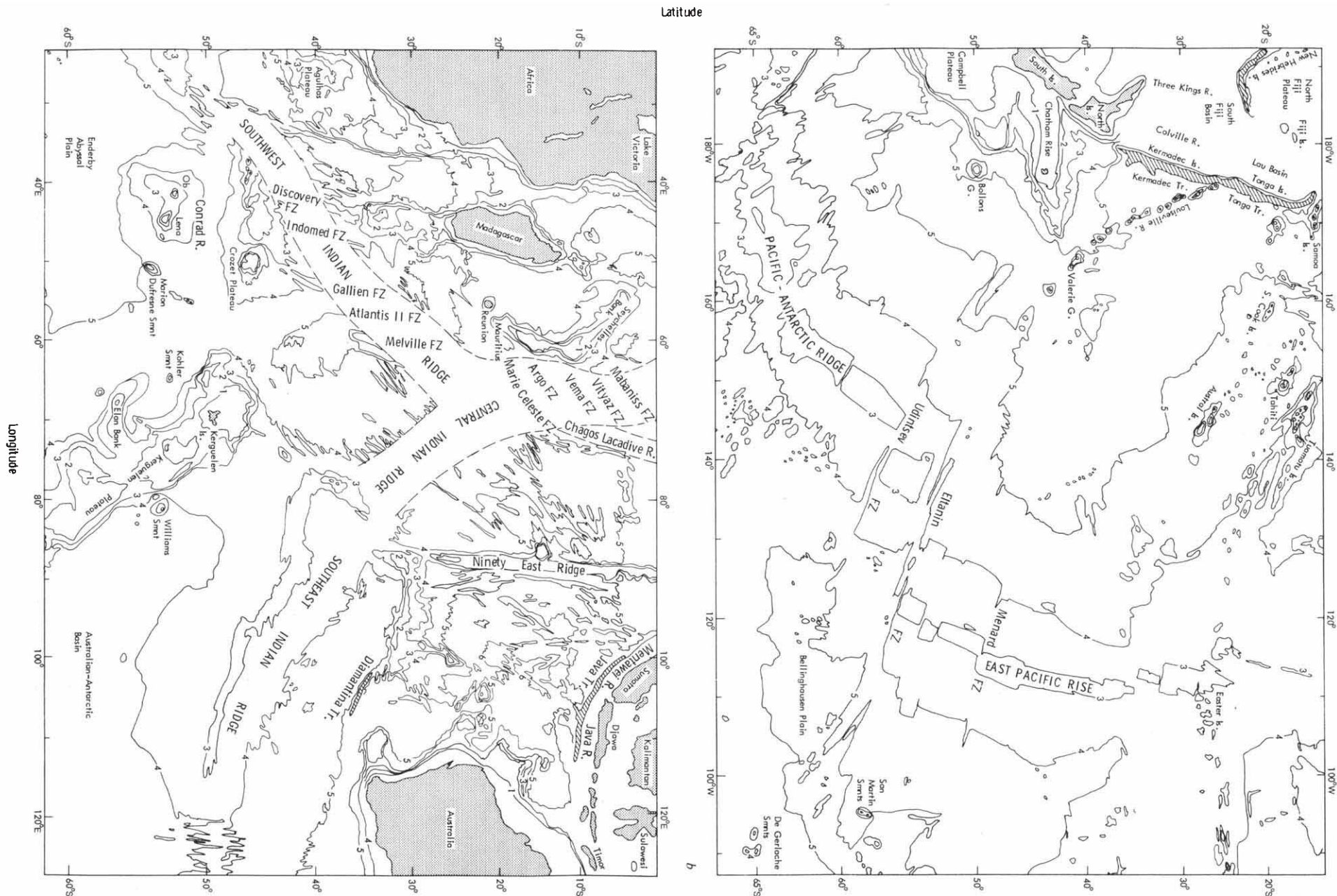


Fig. 2 Existing bathymetric data¹⁰ redrawn to correspond to the regions in Fig. 1. Contour interval, 1 km. Land is indicated with stipple. Areas deeper than 6 km are indicated with heavy striping. R, Ridge or rise; FZ, fracture zone; Smnt, seamount. *a*, South Pacific Ocean. Area west of the Tonga-Kermadec trench has significant topographic complexity and is not contoured. *b*, South Indian Ocean. Area north-east of Java Trench is not contoured. Active spreading centres between dashed lines (South-West and Central Indian Ridges) generally have rough topography shallower than 4 km and are not contoured.

17). In the following discussion, we assume that observed geoid anomalies indicate bathymetric anomalies if ship tracks are inadequate to resolve the postulated feature. On the other hand, some bathymetric anomalies probably exist in our study area with no corresponding geoid anomaly. This represents the effect of a spatially variable compensation mechanism. Gravitational compensation in the oceans is regional, and can be modelled by a thin elastic plate (roughly equivalent to the lithosphere) which overlies a more fluid asthenosphere^{18,19}. The rigid plate supports loads by flexure, and displaced, higher density asthenosphere provides buoyancy. Compensation is thus spread out over an area considerably larger than the topographic feature. The geoid response to a load and its compensation depends on the amount of plate flexure, and is larger if the plate is stiff. Flexural rigidity and elastic thickness are measures of plate stiffness, and increase with age as the lithosphere cools and thickens²⁰. The geoid response to a topographic load therefore depends on the age of the crust at the time of loading. Features formed on young oceanic crust, when effective flexural rigidity is low, will have a correspondingly low geoid response. By about 40–50 Myr, however, elastic thickness and flexural rigidity have increased considerably relative to newly created lithosphere. Thus in regions of older oceanic crust, the geoid response will be more sensitive to young features relative to older ones.

Figure 1 should be interpreted as an indicator of some major bathymetric and/or shallow mass anomalies. The prime utility of these data will be to suggest areas for future detailed ship surveys. The data will be published as transparent overlays by the Scripps Institution of Oceanography, at the same scale as standard GEBCO bathymetric charts¹⁶.

Discussion

Comparison of the geoid anomaly map with existing depth data is best done with bathymetry compiled on transparent overlays at the same scale. When this is done for the North Pacific, most geoid anomalies observed in the Seasat data correspond precisely with known bathymetric features. In contrast, feature-correspondence in the southern oceans is much less exact. We interpret this to mean that bathymetric control is not as good in these poorly surveyed areas. The following discussion focuses on the south-west Pacific and south Indian oceans, where some important discrepancies occur. The geoid anomaly maps (Fig. 1) and bathymetric maps (Fig. 2) have been printed at the same scale so that an overlay can be generated by making a transparency from Fig. 2.

Previously described bathymetric features are clearly visible in Fig. 1a. These include the south-west portion of the East Pacific Rise (EPR), and its extension, the Pacific–Antarctic Ridge (PAR). The large right-lateral offset fracture zones in this spreading system (Eltanin, Udinstev) stand out due to the contrast in crustal age across them.

Along portions of the EPR the geoid high is not symmetrically located with respect to the bathymetrically defined ridge crest. Segments of the EPR between the Menard and Eltanin Fracture Zones have a geoid high some 100–150 km west of the bathymetric high. South of the Udinstev Fracture Zone, the geoid high appears to be located at least 200 km west of the indicated chart position for the ridge. Although the charted ridge positions could be slightly in error, it is also possible that the geoid high is asymmetric with respect to the ridge, possibly related to a locally slow or asymmetric spreading rate or to the history of ridge jumps observed in this region²¹.

Slow spreading centres are characterized by strong topographic relief relative to fast spreading centres. This represents the effect of crustal cooling and subsidence with age²². For a given distance from the ridge, crust formed at a slow spreading centre will be older, and hence deeper than crust formed at a fast spreading centre. The geoid in the vicinity of a spreading centre will tend to show a similar response, due to its correlation

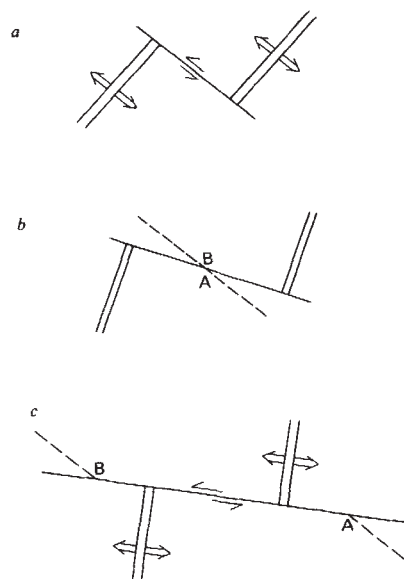


Fig. 3 A shift in the plate rotation pole resulted in a shift of transform fault trends in the Eltanin and Udintsev Fracture Zone systems. Original ridge–transform system (a) experiences instantaneous shift in pole of rotation, with consequent reorganization of ridge and transform fault (b, no spreading has yet occurred). After finite amount of spreading, original transform (dashed line) has been offset (c). The distance A–B is ~2,000 km, or about 35 Myr since the ridge jump in (b) assuming a 6 cm yr⁻¹ total spreading rate. Comparison of the well-defined southeastern bend in the Udintsev Fracture Zone with existing magnetic anomaly maps²³ suggest that this event occurred 45 Myr ago. These estimates compare favourably with better-known North Pacific spreading changes 40–43 Myr ago^{24,25}.

with topography. Hence, the amplitude of a geoid anomaly over a ridge may be inversely correlated with spreading rate. The EPR in the vicinity of the Eltanin and Udintsev fracture zones is characterized by strong positive geoid anomalies (>1 m) relative to the ridge crest north or south. There is also a pronounced geoid high on the long, nearly unbroken east–west PAR segment south of New Zealand. These anomalies are similar in amplitude to those observed over the slow spreading Mid-Atlantic Ridge. Present spreading rates are known to decrease southward along the EPR–PAR system. North of the Menard Fracture Zone (~50°S lat.) total spreading rate is ~10 cm yr⁻¹, decreasing to 6 cm yr⁻¹ south of New Zealand²¹. These rates were also slower in the past²¹. Magnetic anomaly compilations²³ indicate that average total spreading rate between the Menard and Udintsev Fracture Zones since anomaly 16 time (~40 Myr ago) was ~6 cm yr⁻¹.

The Eltanin Fracture Zone consists of two sub-parallel, closely spaced fracture zones (Heezen and Tharp) separated by 100–200 km (ref. 21). The southern one (Tharp) appears to be a direct extension of the Louisville Ridge trend, although it is a broader feature. The geoid high associated with the spreading centre occurs immediately south of the Tharp fracture zone. Parallel fracture zones are not obvious in the geoid data close to the ridge crest, but can be distinguished beyond about 1,000 km from the local spreading axis. The south-east parts of both the Eltanin and Udintsev fracture zones show a distinct change in azimuth, also ~1,000 km from the local axis, implying a change in spreading direction some time in the past. The fracture zones closer to the axis are oriented approximately ESE–WNW, but become nearly SE–NW further away. Molnar *et al.*²¹ describe similar changes in fracture zone trends in the south-west Pacific, and suggested a change in spreading direction between anomaly 20 and anomaly 13 time. The new geoid data allow a better time definition of this event. Figure 3 is a model

for the observed geometry of the Eltanin and Udinstev Fracture Zones. Assuming a uniform 6 cm yr^{-1} spreading rate, the change in geometry occurred about 35 Myr ago. Another estimate can be obtained by noting the location of the well-defined southeastern bend in the Udintsev Fracture Zones ($\sim 127^\circ \text{ W}$ long.), and comparing this with existing magnetic anomaly maps²³. On this basis, spreading direction changed between anomaly 17 and 20 time or ~ 45 Myr ago. These estimates compare favourably with the Eocene change in spreading direction (40–43 Myr ago) inferred from the bend in the Hawaiian–Emperor seamount chain^{24,25}.

Several major seamounts in the Louisville Ridge are apparent in the Seasat data (Fig. 2) which do not appear in recent bathymetry compilations²¹. These are all in regions not covered by ship tracks. Two examples occur at ($\pm 10'$) $42^\circ 00' \text{ S}$ lat., $163^\circ 20' \text{ W}$ long. and $37^\circ 20' \text{ S}$ lat., $169^\circ 20' \text{ W}$ long. In addition, a very large positive feature occurs at $48^\circ 20' \text{ S}$ lat., $153^\circ 00' \text{ W}$ long. The north-west portion of the Louisville Ridge appears to be a closely spaced series of short ($\sim 300 \text{ km}$) linear segments, each consisting of two or three large, closely spaced seamounts. This topographic aspect is clearly different from the Eltanin Fracture Zone to the south-east, and is consistent with a volcanic origin for the Louisville Ridge²⁶.

A major rise or geoid high exists east of the Louisville Ridge, between lat. 38 and 41° S , and long. 160 and 150° W . It is elongated in the east–west dimension, and is $\sim 1,200 \text{ km}$ long by 300 km wide. These dimensions mark it as one of the larger rises of the western Pacific. In detail, it consists of three elevated regions. The westernmost is indicated on Molnar *et al.*'s²¹ compilation, and the western and central ones appear on a recent bathymetry compilation (Fig. 2a). Ship tracks are inadequate to resolve the eastern section indicated by the Seasat data. The westernmost, partially-surveyed region²¹ consists of smaller elevated features rather than a single, broad plateau. They are superimposed on a subtle regional high of no more than 500 m . This topography appears to be inadequate to explain the large observed geoid anomaly, which is $+1.0 \text{ m}$ to $+1.5 \text{ m}$ over most of the area. Either the bathymetric surveys are inadequate, or a significant crustal density anomaly with relatively small corresponding topographic anomaly is implied.

The Indian Ocean geoid anomaly map and corresponding bathymetric chart (Figs 1b and 2b) are dominated by the pattern of spreading ridges, with both geoid and bathymetric highs. The geological evolution of this region has been summarized by MacKenzie and Sclater²⁷. The Central Indian and South-West Indian Ridges are slow spreading centres with strong topographic relief and positive geoid anomalies. In contrast, the South-East Indian Ridge, which continues south of Australia, is a fast-spreading ridge with a small geoid anomaly. A large number of aseismic ridges and plateaus are also evident in the Indian Ocean. Near the south-east coast of Africa, the Agulhas, Mozambique and Madagascar plateaus all have geoid anomalies in excess of $+2 \text{ m}$. A linear, east–west geoid low is evident immediately north of the Diamantina Trough, and extends far to the east, along the southern margin of Australia.

Sea floor topography south of Kerguelen Plateau is very poorly constrained by available bathymetric data. The new bathymetric features which are postulated here and summarized below, occur in regions where ship tracks are sparse or non-existent.

- (1) Two small positive features occur south and south-east of Elan Bank, between about 59 and 61° S lat. The westernmost ($\sim 67^\circ \text{ E}$ long.) is a broad, gentle rise with a geoid anomaly between $+0.5$ and $+1.0 \text{ m}$. The easternmost ($\sim 75^\circ \text{ E}$ long.) is a smaller, more sharply defined seamount with a positive geoid anomaly of at least 2.0 m .
- (2) A large trough east of Kerguelen Plateau is well defined on the geoid map between 54 and 59° S lat.
- (3) A small plateau occurs between Crozet and Kerguelen Islands, centred on 45° S lat., 56° E long.
- (4) Several large positive features occur south of Crozet Island. The largest of these (Conrad Rise) has been previously described, and includes Ob and Lena seamounts. However, this feature appears to be larger than indicated on existing charts, and is different in shape.
- (5) A large seamount exists immediately north-northwest of Marion Dufresne Seamount, adjacent to Conrad Rise. The peak of Marion Dufresne Seamount, previously defined on the basis of a single ship track, lies some 100 km east of its charted position.

Conclusions

Seasat altimeter data can be combined and filtered to provide bathymetric estimates for the poorly surveyed southern oceans. While the prime utility of these data will be for comparison with existing charts to suggest areas for future ship surveys, we suggest some tentative revisions to the existing bathymetric data base for the southern oceans:

- (1) The Louisville Ridge is a nearly continuous feature composed of short ridge segments. The topographic aspect of this feature is clearly distinct from fracture zone topography and suggests a volcanic origin.
- (2) A major east–west elongate plateau or positive gravity anomaly exists east of the Louisville Ridge.
- (3) The south-east parts of the Eltanin and Udintsev Fracture Zone systems contain a well-defined 'hook', consistent with a rearrangement of spreading direction between 35 and 45 Myr ago.
- (4) The Conrad Rise in the south Indian Ocean is larger, and has a different shape than existing charts indicate.
- (5) Several smaller rises or plateaus, and one large seamount (north-northwest of Marion Dufresne) have been identified in the south Indian Ocean.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. We thank Kevin Hussey for image processing and S. M. Smith, J. Mammerrickx, M. McNutt, D. Sandwell and C. Bowin for valuable discussions.

Received 24 January; accepted 17 June 1983.

1. Chelton, D. B., Hussey K. J. & Parke, M. E. *Nature* **294**, 529–532 (1981).
2. Lipa, B. J. & Barrick, D. E. *J. geophys. Res.* **86**, 10921–10930 (1981).
3. Fedor, L. S. & Brown, G. S. *J. geophys. Res.* **87**, 3254–3260 (1982).
4. Tapley, B. D., Born, G. H. & Parke, M. E. *J. geophys. Res.* **87**, 3179–3188 (1982).
5. Talwani, M., Poppe, H. R. & Rabinowitz, P. D. *NOAA Tech. Rep. ETL-228-AMOL* 7–2, pt. 2, 1–34 (1972).
6. McKenzie, D. & Bowin, C. *J. geophys. Res.* **81**, 1903–1915 (1976).
7. McNutt, M. K. *J. geophys. Res.* **84**, 7589–7597 (1979).
8. Watts, A. B. *J. geophys. Res.* **84**, 3817–3826 (1979).
9. Dixon, T. H., Naraghi, M., McNutt, M. K. & Smith, S. M. *J. geophys. Res.* **88**, 1563–1571 (1983).
10. Marsh, J. G. & Williamson, R. G. *J. astr. Sci.* **27**, 345–369 (1980).
11. Tapley, B. D. & Born, G. H. *J. astr. Sci.* **27**, 315–326 (1980).
12. Parke, M. E. & Stavert, L. R. *J. geophys. Res.* (in the press).
13. Sandwell, D. T. *EOS* **63**, 1120 (1982).

14. Lerch, F. J., Marsh, J. G., Klosko, S. M. & Williamson, R. G. *J. geophys. Res.* **87**, 3281–3296 (1982).
15. Marsh, J. G. & Martin, T. V. *J. geophys. Res.* **87**, 3269–3280 (1982).
16. *General Bathymetric Chart of Oceans (GEBCO)* scale 1:10,000,000 (Canadian Hydrographic Service, Ottawa, 1982).
17. Char-Shine Liu, Sandwell, D. T. & Curran, J. R. *J. geophys. Res.* **87**, 7673–7686 (1982).
18. Walcott, R. I. *J. geophys. Res.* **75**, 3941–3954 (1970).
19. McNutt, M. K. & Menard, H. W. *J. geophys. Res.* **83**, 1206–1212 (1978).
20. Watts, A. B., Bodine, J. H. & Ribe, N. M. *Nature* **283**, 532–537 (1980).
21. Molnar, P., Atwater, T., Mammerrickx, J. & Smith, S. M. *Geophys. J. R. astr. Soc.* **40**, 383–420 (1975).
22. Sclater, J. G., Anderson, R. N. & Bell, M. L. *J. geophys. Res.* **76**, 7888–7915 (1971).
23. *Plate-tectonic Map of the Circum Pacific Region, Southeast Quadrant*, scale 1:10,000,000 (American Association of Petroleum Geologists, 1981).
24. Morgan, W. J. *Geol. Soc. Am. Mem.* **132**, 7–22 (1972).
25. Clague, D. A., Dalrymple, G. B. & Moberly, R. *Bull. geol. Soc. Am.* **86**, 991–998 (1975).
26. Hawkins, J. W. *EOS* **54**, 1212 (1973).
27. MacKenzie, D. & Sclater, J. G. *Geophys. J. R. astr. Soc.* **24**, 437–578 (1971).

Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*

Martin Kreitman

Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA

The sequencing of eleven cloned Drosophila melanogaster alcohol dehydrogenase (Adh) genes from five natural populations has revealed a large number of previously hidden polymorphisms. Only one of the 43 polymorphisms results in an amino acid change, the one responsible for the two electrophoretic variants (fast, Adh-f, and slow, Adh-s) found in nearly all natural populations. The implication is that most amino acid changes in Adh would be selectively deleterious.

MANY issues in population genetics require, for proper evaluation, measurements of genetic variation in natural populations. In recent years, this measurement has been made primarily by gel electrophoresis^{1,2}, which detects differences in amino acid sequence of proteins. Such studies have uncovered a remarkable amount of genetic polymorphism at loci coding for soluble enzymes³. However, because bands on electrophoretic gels are phenotypes, not genotypes, the total extent of genetic variation at structural loci remains unknown.

Recent advances in recombinant DNA technology, including rapid sequencing techniques^{4,5}, now allows allelic differences between individual genes to be identified directly at the nucleotide level. The complete resolution of genetic variation afforded by DNA sequence comparison not only solves the problem of detecting all amino acid substitutions between proteins, it also identifies nucleotide variation that is not translated into protein differences. Sequence variation in noncoding regions such as introns, or between synonymous codons, which is presumably not subject to natural selection acting through the phenotype of the encoded protein, is the best available source for estimating the frequency of polymorphism due to neutral gene substitution. This estimate, which can only be made from direct DNA sequence comparison, is crucial for understanding the role of natural selection and genetic drift in the maintenance of genetic variation.

I report here a DNA sequence comparison of 11 independently cloned *Adh* genes from five geographically distinct populations of the fruit fly, *Drosophila melanogaster*. Virtually all natural populations in this species are polymorphic for two electrophoretically distinguishable alleles^{6,7}, *Adh-s* and *Adh-f*, which differ by a single amino acid (Thr versus Lys at codon 192)^{8,9}. However, several other variants have been identified, some of which are electrophoretically indistinguishable from the *Adh-f* or *Adh-s* electromorphs. The most common one, a heat-resistant variant of *Adh-f*¹⁰, has been reported at frequencies as high as 10% in some populations^{11,12}. Therefore, I was particularly interested in whether a DNA sequence analysis could disclose genetic heterogeneity among proteins with identical electrophoretic mobilities. This comparative DNA sequence study of five *Adh-f* and six *Adh-s* alleles, shows the presence of many silent polymorphisms in exons and introns but no amino acid replacement polymorphism within the two electromorphs. This distribution of nucleotide polymorphism is a strong indication that virtually all amino acid replacement mutations have been selectively deleterious. This study also reports extensive sequence conservation in the 3' flanking region of the *Adh* gene.

Physical organization

Figure 1 shows the structure of the *Adh* gene, the protein-encoding sequence of which is split into three sections (of

lengths 96 base pairs (bp), 405 bp and 264 bp) separated by two small introns (65 bp and 70 bp). Two transcripts are produced, a larval and an adult form, differing only in their 5' noncoding leader sequences. The larval mRNA leader sequence, 70 nucleotides in length, is contiguous with the first coding sequence in exon 2 and therefore is unspliced in the mature message. The adult mRNA contains an 87 nucleotide leader sequence that is transcribed from a segment located 690 bp upstream from the first coding sequence. This leader sequence (positions 1–87) is spliced to a site 36 nucleotides 5' to the AUG initiation codon in the mature mRNA (position 741/742). Therefore, in addition to the two small introns dividing the coding regions, the adult primary transcript contains an additional intron of 654 bp at the 5' end of the gene. A larval consensus promoter sequence 'TATAATA' as well as part of the larval leader sequence is included in the 3' end of the adult intron.

Sequence polymorphism

A consensus DNA sequence based on the six *Adh-s* genes is shown in Fig. 2. It contains the complete structural gene, and also flanking 5' (63 bp) and 3' (800 bp) sequence. Table 1 shows a summary of DNA sequence variation among the 11 genes over this 2,721 bp region. Forty-three nucleotide positions are polymorphic overall, 14 in the three coding regions (765 bp), 11 in the adult intron (654 bp), 7 in introns 2 and 3 (135 bp), 3 in the 5' and 3' nontranslated sequences (332 bp) and 8 in the 5' and 3' flanking regions (863 bp). Thirteen of the 43 sites

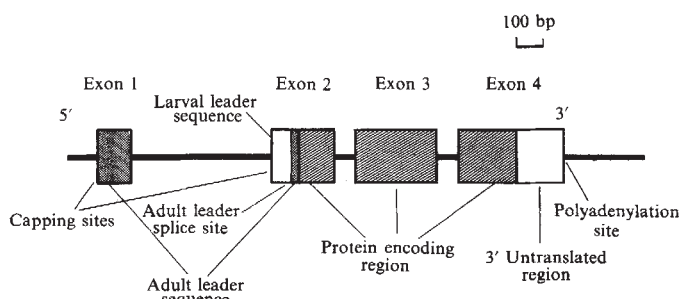


Fig. 1 Structure of the *Adh* gene of *D. melanogaster*. Two different transcripts have been identified in adult and larval tissues²⁶ which differ only in their 5' noncoding sequences. The larval 5' leader is contiguous with the first coding region whereas 87 bp of the adult 5' leader is transcribed from a region located 690 bp further upstream and is spliced to a site located 36 bp upstream from the translation initiation. The two leader sequences overlap, therefore, by 36 bp.

Table 1

Reference sequence	5' Flanking sequence	Adult leader (exon 1)	Intron 1 (Adult intron, larval non-coding)	Larval leader	Translated region of exon 2	Intron 2	Exon 3	Intron 3	Translated region of exon 4	3'-Untranslated region	3' Flanking sequence
	C C G		C A A T A T G G G V1 C V2 G	C	T A C		C C C C	G G A A T	C T C C A* C T A G	A V3 C	A G C V4 C V5 T Δ6
Strain											
Wa-S	. . .		 A T	.	.	.	T T . A	C A . T A	A C	. . .	 Δ
Fl-1S	. . C			.	.	.	T T . A	C A . T A	A C	. . .	 Δ
Af-S	. . .			.	.	.			 A	. . .	. . T V . 1 A .
Fr-S	. . .			.	G T	.			 A	-1 .	T A
Fl-2s	. . .		A G . . . A . T C	A	G G T	.				C 3 .	
Ja-S	. . C			.	G . .	.			. . . T . T . C A	C 4 .	 T . . .
Fl-P	. . C			.	G . .	.			. . G T C T C C .	C 4 .	
Fr-P	T G C		A G . . . A . T C V G V .	.	G . .	.			. . G T C T C C .	C 4 G	
Wa-P	T G C		A G . . . A . T C V G V .	.	G . .	.			. . G T C T C C .	C 4 G	
Af-P	T G C		A G . . . A . T C V G V .	.	G . .	.			. . G T C T C C .	C 5 G	
Ja-P	T G C		A G G G G A . . . V . . T	.	G . .	.	. . A . .	. . G . .	. . G T C T C C .	C 4 .	 -1 . .
No. of polymorphic sites	3	0	11	1	1	2	4	5	9	2	5
Average no. of Nucleotides compared	63	87	620	70	99	65	405	70	264	178	767
% Sites polymorphic	4.7	0	1.8	1.4	1.0	3.1	1.0	7.1	3.5	1.1	0.6

One *Adh-f*, and either one or two *Adh-s* electrophoretic alleles were randomly chosen from isochromosomal lines derived from each of five population samples. S, *Adh-s* alleles; F, *Adh-f* alleles. Collection sites and year collected: Fl, West Palm Beach, Florida, 1979; Wa, Seattle, Washington, 1979; Af, Burundi, Africa, 1977; Fr, Bully, France, 1977; Is, Ishigaki, Japan, 1978. The reference nucleotide sequence is the most common *Adh-s* nucleotide at each of the polymorphic sites. Differences are shown in the body of the table. ∇/Δ: insertion/deletion polymorphisms. The numbers in columns ∇3 and ∇5 are the differences in homopolynucleotide run lengths compared with the consensus sequence. *: Thr-Lys amino acid replacement polymorphism. All other polymorphisms are either silent or noncoding.

Table 2 Summary of silent polymorphism frequencies

	Intron 1	Introns 2+3	Translated region of exon 2	Exon 3	Translated region of exon 4	Complete translated sequence	Non-translated	3' Non-transcribed
Average no. nucleotides compared	654	135	99	405	264	765	335	767
Effective no. silent sites	654	135	25.5	103.6	63.1	192.2	335	767
No. silent polymorphisms	11	7	1	4	8	13	3	5
Per cent silent polymorphism	1.7	5.2	3.9	3.9	14.3	6.7	0.9	0.6

The proportion of nucleotide changes that would be silent among all nine possible substitutions at each codon is averaged over all codons for an *Adh-s* allele, excluding the initiation and stop codon, and excluding any substitution that changes a codon to a stop codon. The effective number of silent sites is calculated for each exon as the product of the total number of nucleotides and the average proportion of all possible substitutions that are silent.

have nucleotide differences that are represented only once in the 11 strains, 11 of which occur in one strain, *Ja-f*. The adult intron and the 3' flanking regions each contain five unique polymorphisms. Of the 14 polymorphic sites in the three coding regions, only one is a polymorphism leading to an amino acid replacement, and it accounts for the two electromorphs. The remaining 13 exon polymorphisms represent silent changes.

Adjacent polymorphisms at the first three non-transcribed nucleotide positions immediately 5' to the adult mRNA cap site (positions -1, -2 and -3 in Fig. 1) represent the only example of consecutively varying nucleotide sites. Polymorphism at these positions strongly suggests that their conservation is not necessary for proper initiation of transcription as all of the fly strains used in this study produce *Adh* (all show *Adh* bands on native acrylamide gels). Although this pattern of polymorphisms is suggestive of a mutational 'hotspot', these same positions are conserved between the commonest sequence in *D. melanogaster* and two sibling species, *Drosophila simulans* and *Drosophila mauritiana*¹³. In addition, there are no analogous polymorphisms in the corresponding larval 5' non-coding region.

Length polymorphism

In addition to nucleotide site variation, the adult intron and the 3' flanking region have six sites containing length polymorphisms, ranging in size from 1 to 37 bp (see Fig. 2). Five of the six polymorphisms involve direct repeats. Two length polymorphisms are structurally similar, involving expansions or contractions of homonucleotide repeats (runs), at position 1,698-1,708 and at position 2,303-2,311. Five different 'run' lengths are associated with the former site, ranging from 10 to 16 bp in length, and three with the latter site, ranging from 9 to 11 bp. The fact that the relationship of lengths among the 11 genes conforms well to the overall pattern of relationship based on shared nucleotide polymorphisms (Table 1), strongly suggests that this length variation is not a cloning artefact. Thus, these homonucleotide runs appear to be mutational hotspots.

The DNA sequences of the two length polymorphisms in the adult intron are shown in Fig. 2. ∇1, located between positions 447/448 and 476/477, contains a 6 bp direct repeat which is tandemly reiterated four times in the insertion. An identical 6 bp sequence is also located adjacent (3') to the point

Fig. 2 Consensus sequence of the *D. melanogaster Adh* gene and flanking regions. The sequence shown is the consensus for the six *Adh-s* alleles. The sites of all the polymorphisms are indicated above the consensus sequences, including insertions (∇) and deletions (Δ). The sequences of the insertions are given below the consensus sequence, and ∨ indicates the precise point of insertion. The horizontal arrows above the consensus sequence at sites ∇1 and Δ6 delineate the regions deleted in these two polymorphisms. The A ← → C polymorphism at position 1,490 is responsible for the Lys ← → Thr difference between the fast and slow *Adh* alleles. The 5' untranslated leader sequence of the adult mRNA is made up of residues 1-87 and 742-777, and that of the larval mRNA of residues 708-777. The 3' end of both mRNAs is at position 1,858. Strain *Af-s* has not been sequenced 3' to position 2,347.

Methods: Cloning and sequencing strategy: An 11.8 kb *Bgl*II restriction fragment containing the *Adh* gene was isolated by plaque hybridization²⁷ from a λ1059 recombinant *Bgl*II genomic library²⁸⁻³⁰, using as a probe sAC-1, a 4.5 kb *Eco*RI *Adh* genomic subclone in pBR322³¹. A 2.7 kb *Cla*I-*Sal*I fragment was subsequently subcloned³² into pBR322. Overlapping nested deletions were constructed by ligating *Msp*I partial digestion products into the *Cla*I site of pBR322. Sets of 5-7 overlapping deletions were selected³³ and sequenced from the common pBR322 *Eco*RI site using the method of Maxam and Gilbert³⁴. Additional restriction sites were used to confirm sequences and to determine flanking region sequences.

of insertion and therefore may represent a template for the tandem repeat. Because 29 bp are deleted in the genes carrying the insertion, the net increase in the size of the intron due to the insertion is only 5 bp.

∇2 is a 37 bp point insertion at position 550/551. The 5' half of this insertion is a 19 bp tandem repeat of a sequence directly adjacent (3') to the point of insertion. The 3' half of the insertion is a 23 bp inverted complementary repeat of a sequence located further downstream from the point of insertion (positions 576–597) and so can form a large, extremely stable secondary structure in either RNA ($\Delta G = -66.9$)¹⁴ or possibly in DNA as shown in Fig. 3. This secondary structure is related to a less stable one ($\Delta G = -28.9$), also shown in Fig. 3, that can form at the same position in genes lacking the insertion. This coincidence argues for the existence of one of the two secondary structures *in vivo*, as a means of generating the length polymorphism.

Relative rates of nucleotide substitution

Transitions and transversions. There is no evidence for an unequal distribution of nucleotide substitutions among the six possible classes of transitions and transversions for the 43 observed polymorphisms (homogeneity $G = 2.6$, $P > 0.5$). The observed ratio of transitions:transversions is 18:25, which does not differ significantly from expectation assuming equal mutation rates and no differential selection ($G = 1.36$, $P > 0.5$). Therefore, transitions do not appear to be accumulating at a significantly faster rate than transversions at the *Adh* locus.

Introns and exons. To test whether silent polymorphisms in the three exons occur at the same frequency as polymorphisms in introns, I calculated the number of 'effectively silent sites'¹⁵ in the coding regions, as explained in Table 2. Overall, 25% (192 out of 765) of the nucleotide positions in the exons are effectively silent. Therefore the per cent polymorphism in the three coding regions, calculated as the observed number of silent polymorphisms $\times 100$ divided by the effective number of silent sites, is $13/192 \times 100 = 6.7\%$ (no significant difference between the three exons: homogeneity $G = 4.472$, $P > 0.1$). Expressed in terms of heterozygosity, two genes chosen at random from the sample differ, on average, at 2.3% of their silent sites.

There is a significant difference in the per cent silent polymorphism between introns (2.4%) and coding regions (6.7%) ($G = 7.36$, $P < 0.01$). However, this difference is accounted for by the significantly lower value in the adult intron (1.7%) than in the two small introns (5.2%) ($G = 5.17$, $P < 0.025$). This suggests that the larval mRNA leader sequence and adjacent noncoding sequence contained in the adult intron are functionally constrained. In contrast, the per cent polymorphism in the two small introns is not significantly different from that in the exons ($G = 0.33$, $P > 0.5$). Therefore, if there are selective constraints on the nucleotide sequence in the two small introns and at effectively silent sites in exons, they appear to be of roughly the same magnitude.

Only five nucleotide sites are polymorphic in the 3' noncoding region out of an average of 767 sites under comparison. Furthermore, each of the five sites are substituted in only a single strain. Even by the conservative comparison of per cent polymorphism, this region (0.65%) has 10-fold lower silent polymorphism than the coding regions (6.7%) and the two small introns (5.2%). It is also significantly lower than the overall level of polymorphism in the exons (1.8%) which includes both amino acid replacement and non-replacement sites ($G = 4.4$, $P < 0.05$). A similar result has also been reported for a 5' flanking region of the β -globin gene¹⁶, which, in interspecific comparisons, is more highly conserved than the introns. Such conservation in flanking regions of structural genes must imply either strong functional constraints or low mutation rates.

Rather than being limited only to the flanking regions of structural loci, this conservation may be representative of the genome in general. Langley *et al.*¹⁷ have obtained an estimate of 0.004 for the heterozygosity per nucleotide site over a 12 kb

region around the *Adh* locus in *D. melanogaster* based on restriction map variation in 18 genomes from four populations. Although this large region may contain other structural loci in addition to *Adh*, their estimate is approximately the same as the average heterozygosity per nucleotide site in exons in this study (0.006). Therefore, large regions of the genome appear to be evolutionarily constrained.

Natural selection against replacement substitutions. The most striking result of the sequence analysis is the extent to which silent polymorphisms (13) outnumber amino acid replacement polymorphisms (1) in the coding region of the *Adh* gene (homogeneity $G = 29.4$, $P < 0.001$). Although it is possible that their underlying rates of mutation are different, it seems more reasonable to assume that the observed numbers reflect differences in the intensity of natural selection acting on the two classes of mutations. Under this assumption, the expected number of replacement polymorphisms, taking the observed number of silent polymorphisms as an estimate of the expected level, is $(765 - 192) \times 13/192 = 39$ amino acid replacements. The large discrepancy between the expected (39) and observed (1) number of replacement polymorphisms clearly indicates that the overwhelming majority of amino acids in the *Adh* polypeptide sequence are constrained by natural selection.

Even if replacement substitutions have been uniformly selected against, this does not imply that the magnitude of selective disadvantage of such mutations has been large. The effectiveness of natural selection at removing deleterious mutations depends on the effective population size, N_e , as well as the magnitude of selection against a mutant allele, s . For example, Kimura defines as 'effectively' neutral¹⁸ only those alleles having selection coefficients, $s < 1/(2N_e)$. When population sizes are large, the fate of even very slightly deleterious mutations will be governed by natural selection.

That effective population sizes are large is suggested by an estimate of the parameter $\theta = 4N_e\mu$, where μ is the amino acid replacement rate per gene per generation. If the 11 *Adh* genes in this study were a random sample drawn from one population, at evolutionary equilibrium the number of silent polymorphisms in the exons would provide a minimum estimate of this parameter. The effect of using a species-wide sample is to overestimate this parameter to the extent that there is population subdivision within the species. However, most of the variation observed in this study is between *Adh-f* and *Adh-s* alleles within, rather than between, populations, implying that most of the species-wide polymorphisms are also present within

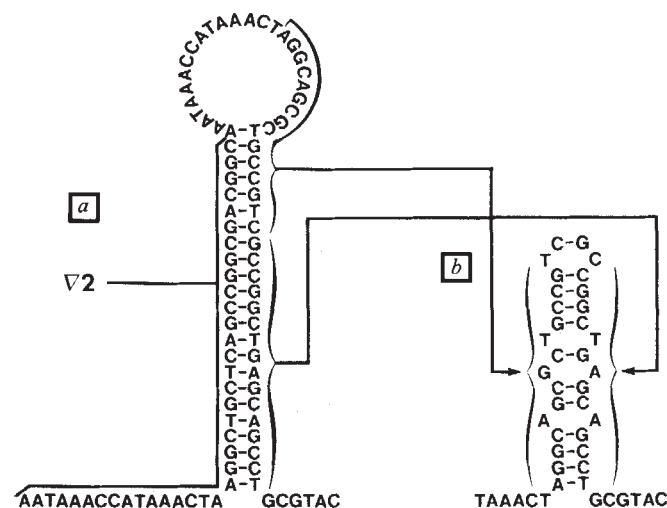


Fig. 3 *a*, Secondary structure predicted from the base sequence involving insertion ∇2. The 16 bp sequence at the 5' end of the insertion that is not involved in the secondary structure is part of a 19 bp tandem repeat of the first 19 bp in the loop. *b*, A less stable secondary structure can also be formed between the two bracketed regions which is 3' to the point of insertion. The 3' locations of the two secondary structures are identical.

populations. Furthermore, in the one population in which the same electromorph is sampled twice (*Adh-s* in Southern Florida), the two genes differ in 6 of 13 possible silent polymorphisms. This suggests that the bias in the estimate of θ resulting from a species-wide sample will not be large. Of course, if silent polymorphisms are themselves selectively constrained, then θ may actually be underestimated.

Silent polymorphisms occur at approximately the same frequency in the coding region and the two small introns (see previous section): I assume this to be, as a first approximation, the equilibrium frequency of neutral polymorphisms. According to the infinite allele model of Kimura¹⁹, the equilibrium number of segregating nucleotide sites at a locus, k , depends only on the effective population size, N_e , and the mutation rate to neutral variants per gene per generation, μ . Then for a sample size of r genes, an unbiased estimator of the parameter $\theta = 4N_e\mu$, is²⁰

$$\hat{\theta} = k/S, \text{ where } S = (1/1) + (1/2) + \dots + (1/(r-1))$$

Setting $k = 13$, the number of segregating silent sites observed in the sample of $r = 11$, then $4N_e\mu = 4.44$. In this case μ_s is the mutation rate at the 192 silent sites out of the 765 sites in the coding region. The mutation rate for amino acid replacements $\mu_r = 3\mu_s$, and therefore $4N_e\mu_r = 13.32$. Assuming a replacement mutation rate at the *Adh* locus of approximately 10^{-6} per generation, as suggested in a study on the mutation rate of several soluble enzyme loci (including *Adh*) to new electromorphs in *D. melanogaster*^{21,22}, the evolutionarily effective species size is then $N_e = 3.3 \times 10^6$. This estimate is consistent with an effective population size of much greater than 10^4 reported by Mukai and Yamaguchi²³ based on allelism of lethals.

If effective population sizes in *D. melanogaster* are reasonably close to the species estimate, the evolutionary fate of even slightly deleterious mutations at the *Adh* locus will be governed primarily by purifying selection rather than genetic drift. Therefore, the high degree of conservation of the *Adh* protein may be the result of natural selection being able to operate on mutants with small deleterious fitness effects.

Given the high degree of conservation within the species, it is of interest that *Adh* in *D. melanogaster* differs by at least three amino acid replacement substitutions¹⁵ from its sibling species, *D. simulans*. Although this difference can be explained by the accumulation of selectively neutral substitutions, it is also consistent with the view that some of these differences have been selectively favoured. This argument rests on the theoretical result that when the effective population size is large, the probability of fixation of an advantageous mutation, being approximately twice its selective advantage²⁴, can be much greater than the probability of fixation of a neutral mutation, which remains approximately $1/(2N_e)$. Therefore, even if selectively favoured mutations arise at the *Adh* locus at a lower frequency than 'effectively' neutral mutations, their relative rate of evolution could be higher.

Phylogenetic comparison of *Adh* alleles

Several additional features of the evolution of *Adh* can be inferred from a different type of analysis by comparing the distribution of shared nucleotides among pairs or groups of alleles: this is a phylogenetic comparison of DNA sequences.

The 11 *Adh* sequences are arranged in Table 1 so that those with the largest number of shared nucleotides at the polymorphic sites are placed next to one another. For example *F1-1S* and *Wa-S* share nine nucleotides not present in any of the other nine alleles, but differ at only three polymorphic sites. This preponderance of shared substitutions implies that these two alleles are much more closely related to each other than to the remaining alleles in the sample. In contrast, the two most distantly related alleles, *Wa-s* and *Ja-f*, share no unique polymorphisms and differ at 33 sites. The alignment of genes in Table 1 indicates three distinct ancestrally related groups, with *F1-1S* and *Wa-S* forming one group, *Af-S* and *Fr-S* a second

group, and the five *Adh-f* alleles a third group. The remaining two genes, *F1-2s* and *Is-S*, are phylogenetically intermediate. The fact that closely related alleles are broadly distributed geographically provides additional evidence that the effective population size is large.

The *Adh-f* alleles can be distinguished from *Adh-s* alleles by only three of the 49 possible polymorphisms, but as a group are considerably more homogeneous than the *Adh-s* alleles. For example, the six *Adh-s* alleles differ from one another, on average, at 15.5 sites, while *Adh-f* alleles differ by 8.1 sites ($F = 1.4173$, $P < 0.025$). There are two possible explanations for this result. One is that the slow allele is ancestral to the fast and has had more time to accumulate nucleotide variation. The alternative assumes that the two alleles are at equilibrium between mutation and drift but that the frequency of the fast allele has been lower, on average, than that of the slow allele. However, the fact that two sibling species, *D. simulans* and *D. mauritiana*, carry the slow allele substitution at codon 192¹³ suggests that *Adh-f* is derived from an *Adh-s* allele.

Although the polymorphic differences distinguishing the 11 strains are clustered along the DNA, as would be expected with tight linkage, two independent intragenic recombination events can be inferred from the distribution of shared polymorphism within both fast and slow alleles. *Fr-s* and *Ja-s* contain the same nucleotide at polymorphic positions 107, 113, 175, 293 and 304 in the adult intron while *F1-2s* and *Fr-f* contain the other nucleotide at each of these sites. This defines two divergent lineages. Yet each lineage is polymorphic for four sites in exon 3—1,452, 1,518, 1,527 and 1,557. Unless these polymorphisms have arisen independently in both lineages, their presence implies a recombination between the 3' and 5' ends of the gene. Similarly, two complementary lineages *F1-1s-Wa-s* and *Af-s-Wa-f*, which can be defined by polymorphisms at positions 1,068, 1,229, 1,283, 1,354, 1,362, 1,400, 1,405 and 1,431, are each polymorphic for two positions in the 5' end of *Adh*, -1 and 175.

The results of this study raise several questions that can only be addressed by additional sequence comparison studies. Having demonstrated a high level of silent nucleotide polymorphism for the *Adh* gene within a species, it is now important to determine how much of this variation occurs in single populations. It is also important to investigate other loci, especially to determine whether the estimate of nucleotide polymorphism and the level of purifying selection observed for *Adh* is representative of other genes. The distribution of polymorphism at loci which are already known to contain many electrophoretic alleles, such as xanthine dehydrogenase and esterase-5 in *Drosophila*, will be particularly revealing: it is possible that intragenic recombination will be important in generating a diversity of alleles from a much smaller number of amino acid replacements²⁵. Finally, the fact that non-coding regions can be very highly conserved relative to introns challenges the assumption that these regions have no functional importance. However, if this conservation is characteristic of most of the genome, its maintenance would be difficult to explain by purifying selection, which would generate very high levels of mutational genetic load in the species. This suggests that mutation rates may vary within the genome.

I thank W. Gilbert, W. Herr, G. Church, D. Strauss, R. Cate, R. Tizard, C. Benyajati, R. Blackman, S. Orzack, L. Brooks, J. Coyne, R. C. Lewontin, T. Mukai, J. David, T. Prout, L. Harley, D. Goldberg and J. Karn for their help. This work was supported by grants GM21179 and GM29301 awarded by the National Institute of General Medical Sciences to R. C. Lewontin.

Received 5 April; accepted 6 May 1983.

1. Hubby, J. L. & Lewontin, R. C. *Genetics* **54**, 577-594 (1966).
2. Lewontin, R. C. & Hubby, J. L. *Genetics* **54**, 595-609 (1966).
3. Coyne, J. A. *Isozymes: Current Topics in Biological & Medical Research* Vol. 6, 1-32 (Liss, New York, 1982).
4. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
5. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
6. Johnson, F. M. & Schaffer, H. E. *Biochem. Genet.* **10**, 149-163 (1973).

7. Oakshott, J. G. *et al. Evolution* **36**, 86–96 (1982).
8. Benyajati, C., Place, A. R., Powers, D. A. & Sofer, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2717–2721 (1981).
9. Retzius, A. D. & Thatcher, D. R. *Biochimie* **61**, 701–704 (1980).
10. Chambers, G. K. *Genetic Studies of Drosophila Populations* (eds Gibson, J. B. & Oakshott, J. G.) 77–94 (Australian National University, Canberra, 1981).
11. Sampsel, B. *Biochem. Genet.* **15**, 971–988 (1977).
12. Wilkes, A. V., Gibson, J. B., Oakshott, J. G. & Chambers, G. K. *Aust. J. biol. Sci.* **33**, 575–585 (1980).
13. Bodmer, M. & Ashburner, M. (personal communication).
14. Borer, P. N., Dengler, B. & Tinoco, I. *J. molec. Biol.* **86**, 843–853 (1974).
15. Holmquist, R., Cantor, C. R. & Jukes, T. H. *J. molec. Biol.* **64**, 145–162 (1972).
16. Moschonas, N., deBoer, E. & Flavell, R. A. *Nucleic Acids Res.* **10**, 2109–2120 (1982).
17. Langley, C. H., Montgomery, E. & Quattlebaum, W. F. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5631–5635 (1982).
18. Kimura, M. *Nature* **217**, 624–626 (1968).
19. Kimura, M. *Genetics* **61**, 893–903 (1969).

20. Ewens, W. J. *Mathematical Population Genetics* (Springer, New York, 1979).
21. Mukai, T. & Cockerham, C. C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2514–2517 (1977).
22. Voelker, R. A., Shaffer, H. E. & Mukai, T. *Genetics* **94**, 961–968 (1980).
23. Mukai, T. & Yamaguchi, O. *Genetics* **76**, 339–366 (1974).
24. Crow, J. F. & Kimura, M. *An Introduction to Population Genetic Theory* (Burgess, Minneapolis, 1970).
25. Strobeck, C. & Morgan, K. *Genetics* **88**, 829–844 (1978).
26. Benyajati, C., Spoerel, N., Haymerle, H. & Ashburner, M. *Cell* **33**, 125–133 (1983).
27. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
28. Maniatis, T. *et al. Cell* **15**, 687–701 (1978).
29. Enquist, L. & Sternberg, N. *Meth. Enzym.* **68**, 281–298 (1979).
30. Karn, J., Brenner, S., Barnett, L. & Cesareni, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5172–5176 (1980).
31. Goldberg, D. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5794–5798 (1980).
32. Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961 (1975).
33. Holmes, D. S. & Quigley, M. *Analyt. Biochem.* **114**, 193–197 (1981).
34. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).

LETTERS TO NATURE

Discovery of a 6.1-ms binary pulsar PSR1953+29

V. Boriakoff*, R. Bucccheri† & F. Fauci‡

* National Astronomy and Ionosphere Center, Cornell University, Ithaca, New York 14850, USA

† Istituto di Fisica Cosmica e Informatica, Consiglio Nazionale delle Ricerche, Palermo, Italy

‡ Istituto di Fisica, GIFCO-CNR, Università di Palermo, Palermo, Italy

A systematic search for fast radio pulsars in some of the error boxes of the γ -ray satellite COS B¹ unidentified point sources^{2,4} on the 305-m Arecibo Observatory radiotelescope was started 3 years ago. The motivation for the search and the method and limitations are given elsewhere⁵. One of the search areas covered was the γ -source 2CG065+00. We report here the discovery of a 6.13317-ms pulsar in a binary system with a 120-day orbital period inside the error box of this source. The initial search observation (23 July 1980) with a sampling rate of 4.167 ms produced an aliased (76.93 Hz) detection at 430 MHz. Confirmatory observations made in March 1983, showed a variable pulse frequency indicating a Doppler shift due to a periodic orbit around another object⁶. A compilation of measured system parameters and the underlying assumptions are given in Table 1.

Figure 1 shows the pulsar pulse frequency-time relationship. Data taken at 430 MHz in the search mode contain simultaneous scanning of the output of 30 contiguous-in-frequency filters. Because the initial confirmation runs were done in the search mode with 0.25-MHz bandwidth per filter (dispersion time in one filter bandwidth is 44% of the pulse period), error in the pulse frequency was relatively large, but this method of data collection was continued until other parameters were determined. Later on dedispersed observations were used with a filter bank of 20 kHz bandwidth per filter with better timing accuracy⁷. At the same time single-pulse arrival timing was started. Because only about half of an orbit has been observed, the orbital parameters determined up to now may have errors. However, initial fittings of a sinusoid to the pulse frequency show a very good approximation, although at present a non-zero orbital eccentricity cannot be excluded. Assuming $e = 0$ a best-fit sinusoid has a period of 120 days and a peak velocity value of 6.02 km s⁻¹. Care was taken to rule out the possibility of an aliased orbital period. The worst case is a 23.74 or a 24.14-h period, all other orbital periods shorter than these will show drastic pulse frequency changes over the maximum possible 2 h of observation. On days when dP/dt due to the Doppler shift was maximum, pulse frequencies close to both ends of the 2 h of observation were compared and found to be consistent only with $P_{\text{orb}} = 120$ days.

The July 1980 observation is close to the lowest pulse frequency of the 1983 half-orbital period observations. Assum-

ing that indeed it is the lowest frequency a maximum value of $\dot{P}$ can be computed by taking the difference between the 1980 and 1983 minima. As the 1980 point is not necessarily the minimum, a smaller value of $\dot{P}$ is likely; this includes negative $\dot{P}$ values. We have assumed the shape and size of the orbit to be invariant over the 3 yr interval. A better value of $\dot{P}$ will be obtained when the orbital parameters are sufficiently well

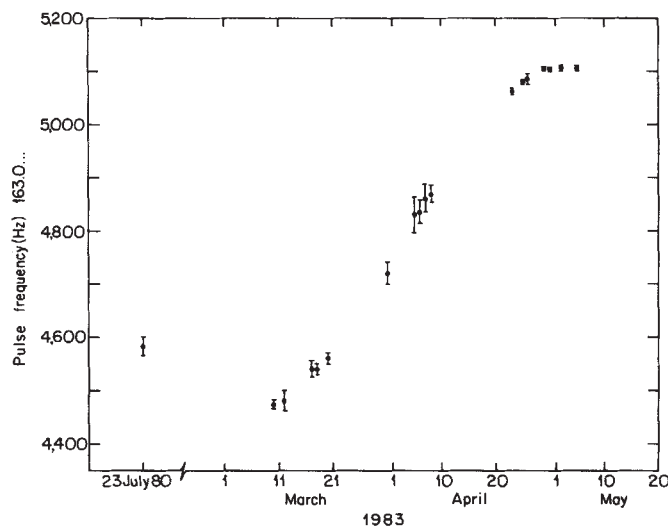


Fig. 1 Frequency of pulses at the Solar System barycentre of PSR1953+29 function of time. Accuracy is dependent on the data taking system used.

Table 1 Preliminary parameters of the millisecond binary pulsar PSR1953+29

VLA position (1950.0):
RA = 19h 53min 26.7s ± 0.24 s
Dec. = 29° 00' 42.0" ± 3.2 "
$l = 65.84^\circ$
$b = +0.443^\circ$
$P = 6.13317 \text{ ms} \pm 0.00002$ on JD = 2445427.96
$\dot{P} \leq 5.8 \times 10^{-16} \text{ s s}^{-1}$
Characteristic age = $P/2\dot{P} \geq 1.675 \times 10^5 \text{ yr}$
$B \leq 6.0 \times 10^{10} \text{ G}$ (assuming $I = 10^{45} \text{ g cm}^2$, $R = 10^6 \text{ cm}$, $\alpha = \pi/2$)
Velocity of light cylinder radius = 292.6 km
Projected orbital velocity peak value = $K_p = 6.02 \text{ km s}^{-1} \pm 0.02$
Dispersion measure = $104.5 \text{ pc cm}^{-3} \pm 0.03$
Distance = 3.5 kpc (assume $\langle n_e \rangle = 0.03 \text{ cm}^{-3}$)
Flux at 1,385 MHz as a continuum source = 1.0 mJy ± 0.5
If the orbit is circular ($e = 0$) then
$P_{\text{orb}} = 120 \text{ days} \pm 4$
$a_p \sin i = 9.9 \times 10^6 \text{ km} \pm 0.1$
Mass function = $\frac{m_c^3 \sin(i)}{(m_p + m_c)^2} = 0.00272 M_\odot$

known to extrapolate accurately to 1980. A lower limit for the characteristic age $P/2\dot{P}$ is given, as well as the upper limit for the magnetic field⁸, where we assumed a neutron star moment of inertia of 10^{45} g cm², a radius of 10^6 cm and a magnetic field orthogonal to the rotation axis.

From the general expression of the radial velocity of a spectroscopic binary and the peak value of radial velocity we determined the $a_p \sin i$ value where a_p is the semimajor axis of the pulsar orbit and i is the angle between the orbital plane and the plane perpendicular to the line-of-sight. Another parameter obtainable with our measured values and the $e = 0$ assumption is the keplerian mass function

$$f(M_\odot) = \frac{4\pi^2 a_p^3 \sin^3(i)}{G P_{\text{orb}}^2} = \frac{m_c^3 \sin^3 i}{(m_p + m_c)^2} = 0.00272 M_\odot$$

where m_c is the companion mass and P_{orb} is the orbital period. Assuming $m_p = 0.5, 1.0, 1.4$ and $2.0 M_\odot$ we can plot the relationship between m_c and $\cos i$ as shown in Fig. 2. It follows that for most i values $m_c < 1 M_\odot$.

Because in our search mode all 30 channels of the filter bank were available for further processing, a value of the dispersion

measure was quickly obtained by crosscorrelating pulse profiles obtained from different channels (different observing frequencies). Assuming an average interstellar electron density $\langle n_e \rangle = 0.03$ cm⁻³ a distance of 3.5 kpc is inferred. Because of the small value of b (0.443°) the addition of a z gradient in n_e changes the result only slightly.

A 45-min integration Very Large Array (VLA) map of the area was quickly made (courtesy of B. Clark of NRAO) at 1,480 MHz and is shown in Fig. 3. The pulsar may have some structure around it, but a substantially longer integration is required to bring out the details. No clearly defined ring structure is apparent.

Figure 4 shows the integrated pulse profiles at 1,385.125 and 430.325 MHz. The difference in structure is striking. Although at 430 MHz only one polarization is used, observations at the other circular polarization show a nearly identical profile, so the figure is a good representation of the intensity pulse profile. It is not clear from these two profiles which 1,385 MHz component matches the strong 430 component, however, observations at 606 MHz (one linear polarization only) seem to show an intermediate profile: a strong central component and two wings at approximately half the strength of the central component. This suggests that the centre peak at 1,385 MHz matches the 430 MHz one. The very weak peak at $\sim 60^\circ$ long. may be an interpulse; integration was not carried out long enough to be absolutely certain.

The most drastic feature setting the 6.1-ms pulsar apart from the 1.5-ms pulsar, is its very large duty cycle. At 1,385 MHz it is of the order of 42% of the pulse period. If the presumed interpulse is included the duty cycle is raised to 52%. This is an anomalous situation even for slow pulsars, where the average duty cycle is $\sim 10\%$. Traditionally in the polar cap model^{9,10}, a large duty cycle is explained by a small angle α between the rotation axis and the magnetic field. The presence of an interpulse, if confirmed, poses a problem: interpulses are traditionally explained as the emission from the polar cap opposite

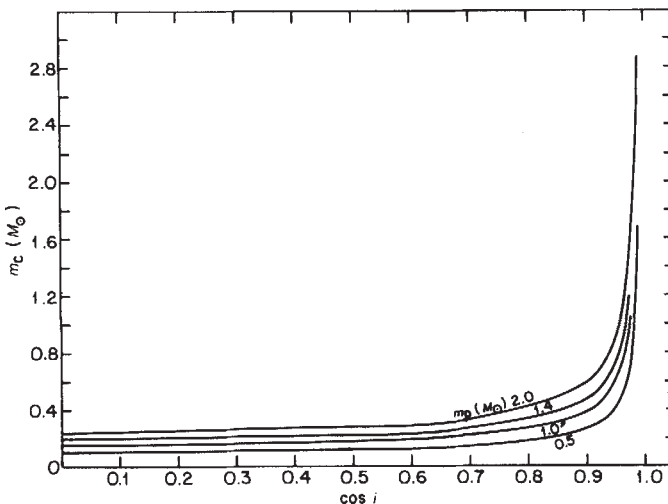


Fig. 2 Relationship between four assumed pulsar masses m_p and the companion mass function of the angle i between the orbital plane and the plane perpendicular to the line-of-sight.

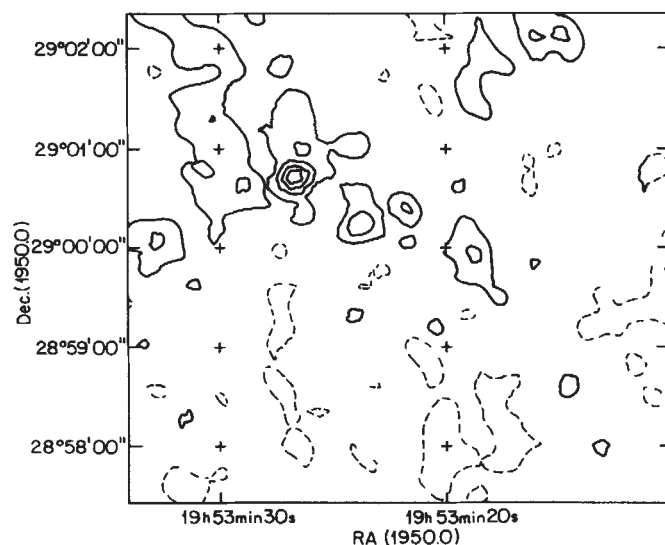


Fig. 3 VLA map of the area around PSR1953+29 done at 1,480 MHz. Each contour is equivalent to 0.3 mJy.

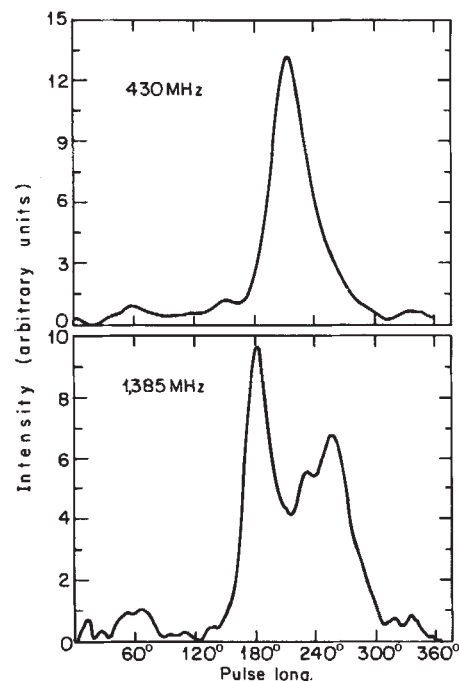


Fig. 4 Average pulse profiles at 1,385.125 MHz (intensity, 7.75 MHz bandwidth) and 430.325 MHz (single circular polarization, 0.62 MHz bandwidth). 98.7% of the pulse period is displayed. No relative phase alignment between the two profiles was done. Signal was dedispersed in a post-detection dedispersion system¹², the time resolution is 260 μ s at 430 MHz (15.2°) and 141 μ s at 1385 MHz (8.3°).

to that of the main pulse, and this requires $\alpha \sim \pi/2$ contradicting our previous requirement. If $\alpha = \pi/2$ and the field is dipolar to obtain the observed duty cycle, we should have radiation generated at $r = 150$ km, that is, about half of the velocity of light radius. Assuming the same polar cap α should be 6.1° to provide the observed duty cycle with radiation generated at $r = 20$ km, a more likely situation (if the polar cap gap is in operation). The interpulse, if generated by the opposite polar cap would be invisible. However, in this case, emission from one distorted main polar cap could provide emission at other longitudes, including the presumed interpulse.

A search of the area for a point source or ring structures in the X-ray images of the Einstein Observatory Image Proportional Counter produced only upper limits⁹. Nevertheless, if the X-ray emission is orbital-phase dependent it may have been missed by the X-ray observation, which lasted < 1 h. It is also possible that, as with the PSR0833-45 (Vela) pulsar, the X-ray emission is absent, although there is γ -ray emission.

Assuming a $\dot{P}$ equal to the determined upper limit we can estimate if a physical association with the γ -ray source is possible. The Crab pulsar total energy loss is $\dot{E}_c = 4.68 \times 10^{38} \text{ erg s}^{-1}$, the pulsed γ -ray energy^{2,11} (assuming a beaming factor of $1/4\pi$) is $\dot{E}_{cy} = 6.88 \times 10^{33} \text{ erg s}^{-1}$ for the energy range of 50 keV to 5 MeV (COS B range). The total energy loss of PSR1953+29 for a $\dot{P} = 0.58 \times 10^{-15} \text{ ss}^{-1}$ and a moment of inertia of 10^{45} g cm^2 is $\dot{E}_{P1953} = 10^{38} \text{ erg s}^{-1}$. If we assume the same γ -ray beaming factor, a distance of 3.5 kpc and ascribe the γ -ray emission² to PSR1953+29 then we find $\dot{E}_{P1953\gamma} = 6.5 \times 10^{33} \text{ erg s}^{-1}$. The ratios

$$\frac{\dot{E}_{cy}}{\dot{E}_c} = 1.47 \times 10^{-5} \text{ and } \frac{\dot{E}_{P1953\gamma}}{\dot{E}_{P1953}} = 6.67 \times 10^{-5}$$

are remarkably close. Proof of the association between 2CG065+00 and PSR1953+29 will only be possible through γ -ray pulsation studies, or by a much more accurate position determination of 2CG065+00.

A quick search for pulses from the companion has provided only an upper limit: it has a peak flux of < 0.1 that of the pulsar at 430 MHz for pulse periods between ~ 2 ms and ~ 8 s.

Because of the low eccentricity of the orbit (perhaps even $e = 0$) there seems to have been an episode of mass transfer in the system's history. This episode must have followed any supernova event, perhaps the one in which the neutron star of the pulsar was formed. The companion may be a low-mass white dwarf, although this is only a probabilistic argument: there is $\leq 20\%$ probability ($m_p = 2M_\odot$) that the mass of the companion is $m_c > 1M_\odot$. Searches for the effects of the companion's atmosphere on the pulsar pulses may resolve this question, as well as optical searches of counterparts.

We thank D. C. Ferguson, R. Isaacman, and other NAIC personnel for assistance in the project. The Arecibo Observatory is part of the National Astronomy and Ionosphere Center which is operated by Cornell University under contract with the NSF. This research has been partially supported by the Italian National Research Council, under the Ufficio Relazioni Internazionali. We thank the Caravane collaboration for stimulating discussions.

Received 29 June; accepted 6 July 1983.

- Scarsi, L. in *Recent Advances in Gamma-Ray Astronomy* (eds Wills, R. D. & Batrick, B.) 3 (European Space Agency, Paris, 1977).
- Swaneburg, B. N. *et al. Astrophys. J. Lett.* **243**, L69 (1981).
- Hermesen, W. *Phil. Trans. R. Soc. A* **301**, 519 (1981).
- Buccheri, R., Morini, M. & Sacco, B. *Phil. Trans. R. Soc. A* **301**, 495 (1981).
- Boriakoff, V., Buccheri, R., Fauci, F. & Ferguson, D. C. *Astrophys. J.* (submitted).
- Boriakoff, V., Buccheri, R. & Fauci, F. *IAU Circ. No.* 3806 (1983).
- Boriakoff, V., Ferguson, D. C., Haugan, M. P., Terzian, Y. & Teukolsky, S. A. *Astrophys. J. Lett.* **261**, L97 (1982).
- Pacini, F. *Nature* **216**, 567 (1967).
- Helfand, D. J., Ruderman, M. A. & Shaham, J. *Nature* **304**, 423-425 (1983).
- Ruderman, M. A. & Sutherland, P. G. *Astrophys. J.* **196**, 51 (1975).
- Buccheri, R. in *Pulsars* (eds Sieber, W. & Wielebinski, R.) 241 (Reidel, Dordrecht 1981).
- Boriakoff, V. *Astr. Astrophys. Suppl.* **15**, 479 (1974).

On the origin of the 6.1-ms pulsar

P. C. Joss & S. A. Rappaport

Department of Physics, Center for Space Research, and Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Boriakoff *et al.*^{1,2} have recently reported the discovery of a pulsar, PSR1953+29, with a pulse period, P , of only 6.1 ms, and with period derivative $\dot{P} < 6 \times 10^{-16} \text{ ss}^{-1}$. This is only the second known pulsar with P in the millisecond range, the first being PSR1937+214^{3,4}. PSR1953+29 is in a binary system with an unseen companion: its orbital period, P_{orb} , is ~ 120 days, its projected semi-major axis, $a_p \sin i$, is $\sim 9 \times 10^{11} \text{ cm}$, and its orbital eccentricity, e , is small. We present here a model for the origin and evolution of this binary system that quantitatively accounts for all of its salient features. Our evolutionary scenario begins with a luminous binary X-ray source (see also refs 17, 18) composed of a neutron star and a lower giant-branch companion (we do not attempt to understand the preceding evolution, which may well have included a common-envelope phase^{5,6}), and terminates when the system has evolved into its present configuration and the companion has become a low-mass degenerate dwarf. We show that our model can explain both the large separation of the binary and the approximate circularity of the orbit despite the large separation relative to the present sizes of the component stars (which tends to render inoperative the tidal dissipation that might otherwise circularize the orbit^{7,8}).

The scenario begins with a lower giant-branch star that is evolving on a nuclear time scale, in orbit with a neutron star. When the size of the giant reaches that of its Roche lobe, mass transfer onto the neutron star commences. If the magnetic dipole moment, μ , of the neutron star at this epoch is sufficiently small ($\mu \ll 10^{30} \text{ G cm}^3$), a source similar to the bright galactic-bulge X-ray sources may result^{9,10}. The mass transfer, once underway, proceeds on the nuclear time scale of the giant companion^{9,10}. When the hydrogen-rich envelope of the giant has been transferred to the neutron star (or lost from the system), the degenerate helium-rich core remains behind.

Table 1 Calculated model binary parameters

Initial binary parameters	PSR1953+29		PSR0820+02
	$Z = 10^{-4}$	$Z = 0.02$	$Z = 0.02$
$M_c (M_\odot)$	1.00	1.00	1.00
$M_p (M_\odot)$	1.30	1.30	1.30
$M_{\text{He}} (M_\odot)$	0.318	0.241	0.380
$R_c (10^{11} \text{ cm})$	11.4	7.2	64
$a (10^{11} \text{ cm})$	33.2	20.7	186
$a_p (10^{11} \text{ cm})$	14.4	9.0	81
$P_{\text{orb}} (\text{days})$	25.2	12.2	335
Final binary parameters			
$M_c (M_\odot)$	0.41	0.31	0.46
$M_p (M_\odot)$	1.89	1.99	1.84
$M_{\text{He}} (M_\odot)$	0.41	0.31	0.46
$R_c (10^{11} \text{ cm})$	24.0	21.9	119
$a (10^{11} \text{ cm})$	92	93	442
$a_p (10^{11} \text{ cm})$	16.5	12.4	87
$P_{\text{orb}} (\text{days})$	117	118	1225
$f(M) (M_\odot)$	0.0131	0.0054	0.0178
$T (\text{yr})$	3.5×10^7	7.7×10^7	3.8×10^6
$i (\text{deg})$	34^{+9}_{-17}	48^{+14}_{-30}	34^{+10}_{-15}

R_c , effective radius of the companion star; a , total orbital separation; other parameters are defined in the text. $f(M) = M_c^3 / (M_c + M_p)^2$ is the mass function divided by $\sin^3 i$. The error bars on i reflect reasonable limits, based on extreme assumptions for the present mass of the neutron star ($0.5 M_\odot \leq M_p \leq 2.7 M_\odot$).

The evolution of a system of this type can be explicitly calculated by use of the radius-core mass relationship for lower-branch giants together with conventional techniques for binary stellar evolution computations (ref. 10 and refs therein). One of us (S.A.R.) recently carried out such calculations in collaboration with Webbink and Savonije¹⁰; we have here adopted and extended these calculations to suit the present problem. For specificity, we assume that the mass transfer is conservative (that all the mass lost by the giant is accreted by the neutron star). The adjustable parameters in this model are the initial

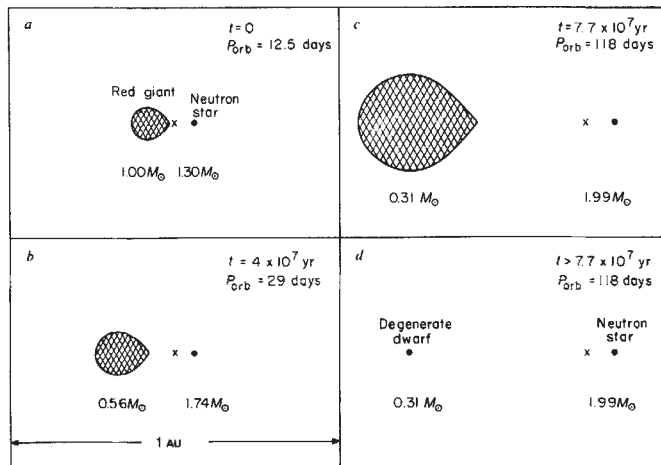


Fig. 1 The evolution of a binary system consisting of a lower-branch giant undergoing mass transfer to a neutron star on a nuclear time scale (drawn to scale). Initially, the giant has a total mass of $1.0 M_{\odot}$, a helium-rich core of mass $0.24 M_{\odot}$, and a surface metallicity of $Z = 0.02$. In each of the four frames, the elapsed time, t , since the onset of mass transfer, the orbital period, P_{orb} , and the masses of the two stars are indicated. Conservative mass transfer has been assumed. $\times$, Location of system centre-of-mass.

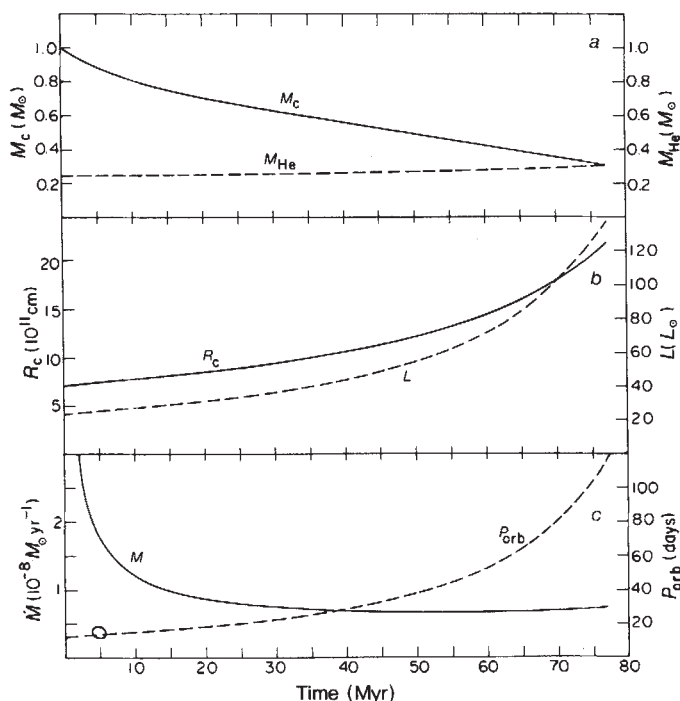


Fig. 2 Evolution of a binary system with a lower giant-branch secondary component of initial mass $1.0 M_{\odot}$ and surface composition $X = 0.70$, $Y = 0.28$, $Z = 0.02$. Plots as functions of time: *a*, total mass (solid line) and core mass (dashed line) of the secondary; *b*, radius (solid line) and intrinsic bolometric luminosity (dashed line) of the secondary; *c*, mass accretion rate onto the neutron star (solid line) and orbital period (dashed line). Conservative mass transfer has been assumed. The initial mass of the neutron star was taken to be $1.3 M_{\odot}$.

values of masses of the neutron star and companion star, M_p and M_c respectively, the initial value of mass, M_{He} , of the helium-rich core of the companion, and the metallicity, Z , of the giant envelope. For simplicity, we chose the initial value of $M_c = 1.0 M_{\odot}$ (although other values are possible) and $M_p = 1.3 M_{\odot}$ (a plausible value for a neutron star when it is first formed¹¹). We must, in fact, have $M_c < (5/6)M_p$ if the mass transfer process is to be both conservative and stable¹⁰. Initial values of $M_p \geq 1.3 M_{\odot}$ are allowable, provided that the final mass of the pulsar, following mass transfer, is less than the maximum stable mass of a neutron star. (Since the companion star fills its Roche lobe throughout the mass transfer process, the orbital period is nearly independent of M_p (see, for example, ref. 12); hence, our choice for M_p is unimportant.) We considered two rather extreme values for Z ($Z = 10^{-4}$ and $Z = 2 \times 10^{-2}$). We then chose M_{He} so that P_{orb} is equal to the observed value at the end of the mass transfer process. Following the termination of mass transfer, the orbital parameters are virtually fixed; the orbit is sufficiently wide that neither tidal dissipation^{7,8} nor gravitational radiation¹³ can effectively alter the orbit. Relaxation of the requirement of conservative mass transfer would allow a wider range of initial binary separations and final masses for the neutron star, but would otherwise not affect our conclusions.

The results of the evolutionary calculations are summarized in Table 1 and illustrated in Figs 1 and 2. The following points are noteworthy.

- (1) Before and during mass transfer, the orbital separation is not very much larger than the size of the companion, so that tidal dissipation will effectively circularize the orbit. Moreover, the elapsed time, T , during mass transfer is long compared with the orbital period, so that the orbital parameters will respond adiabatically: the orbit will expand greatly but remain circular. Thus, both the circularity and the large size of the final orbit are natural consequences of our model.
- (2) The extremely short pulse period probably results from the spin-up of the neutron star by accretion torques (see ref. 14 and references therein). In order to reach such a short rotational period, the magnetic moment of the neutron star should already have been rather small ($\mu \leq 3 \times 10^{27} \text{ G cm}^3$) at the time of mass transfer (see, for example, ref. 17). This conclusion is consistent¹⁵ with the rather small upper limit on $\dot{P}$ for this pulsar.
- (3) We find that for a wide range of initial conditions the present value of M_c should lie in the range ~ 0.31 – $0.41 M_{\odot}$ and the orbital inclination angle, i (as inferred by comparison of the observed and model mass functions), should lie in the range of $\sim 20^\circ$ – 60° . This range of possible values for the mass of the companion is the most secure prediction of the model, since the final value of M_c fixes the final orbital period for a given value of Z . In fact, an accurate measurement of M_c would not only help to confirm the validity of this model, but would also provide an interesting check on our understanding of the structure of lower-branch giants.
- (4) The degenerate-dwarf companion should be intrinsically very faint (absolute bolometric magnitude ≥ 11 for effective temperatures $\leq 10^4 \text{ K}$), unless the mass transfer terminated only recently. This conclusion is consistent with the lack of a conspicuous optical counterpart, brighter than ~ 18 mag, to PSR1953+29 on the Palomar Sky Survey prints. Moreover, there should no longer be any observable blast-wave supernova remnant associated with this source, although the existence of a faint nonthermal nebulosity, maintained by the low-power output² of the pulsar, cannot be theoretically excluded.

The same type of model can be invoked for one of the remaining three known binary pulsars, PSR0820+02 ($P \approx 0.86 \text{ s}$; $\dot{P} \approx 1.2 \times 10^{-16} \text{ ss}^{-1}$; $P_{\text{orb}} \approx 1,232 \text{ days}$; $a_p \sin i \approx 4.9 \times 10^{12} \text{ cm}$; $e \approx 0.012$; ref. 16). An illustrative evolutionary model calculation, leading to orbital elements consistent with the above parameters, is given in Table 1. The most striking difference between this system and PSR1953+29 is the much longer pulse period. This may indicate a much larger neutron-star magnetic moment during the mass-transfer epoch ($\mu \geq$

10^{30} G cm^3) and suggests that during this epoch the system may have been visible as an X-ray pulsar rather than a bright galactic-bulge X-ray source.

This research was supported in part by the NSF under Grant AST 82-17451 and by NASA under grants NSG 7643 and NGL 22-009-638 and contract NAS5-24441.

Received 27 June; accepted 12 July 1983.

1. Boriakoff, V., Buccheri, R. & Fauci, F. *IAU Circ. No. 3806* (1983).
2. Boriakoff, V., Buccheri, R. & Fauci, F. *Nature* **304**, 417–419 (1983).
3. Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. *IAU Circ. No. 3743* (1982).
4. Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. *Nature* **300**, 615 (1982).
5. Paczyński, B. *Structure and Evolution of Close Binary Systems*, IAU Symp. No. 88, 75 (1976).
6. Meyer, F. & Meyer-Hofmeister, E. *Astr. Astrophys.* **78**, 167 (1979).
7. Lecar, M., Wheeler, J. C. & McKee, C. F. *Astrophys. J.* **205**, 556 (1976).
8. Zahn, J.-P. *Astr. Astrophys.* **57**, 583 (1977); erratum **67**, 162 (1978).
9. Taam, R. E. *Astrophys. J.* (in the press).
10. Webbink, R. F., Rappaport, S. & Savonije, G. J. *Astrophys. J.* (in the press).
11. Joss, P. C. & Rappaport, S. *Nature* **264**, 219 (1976).
12. Joss, P. C., Avni, Y. & Rappaport, S. *Astrophys. J.* **221**, 645 (1983).
13. Landau, L. D. & Lifshitz, E. M., *The Classical Theory of Fields* 2nd edn (Pergamon, Oxford, 1962).
14. Ruderman, M. A. & Shaham, J. *Comments Astrophys. Space Sci.* (submitted).
15. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).
16. Manchester, R. N. *et al. Astrophys. J.* **268**, 832 (1983).
17. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728 (1982).
18. Fabian, A. C., Pringle, J. E., Verbunt, F. & Wade, R. A. *Nature* **301**, 222 (1983).

Evolutionary history of the 6.1-ms pulsar

Bohdan Paczynski*

Princeton University Observatory, Princeton,
New Jersey 08544, USA

Boriakoff *et al.*¹ have reported the discovery of a 6.1-ms radio pulsar in a binary with an orbital period of 120 days, $a_1 \sin i = 0.99 \times 10^{12} \text{ cm}$, and $e \approx 0$. This is the fourth binary radio pulsar, and the third to have a nearly circular orbit². The origin of such systems has been discussed frequently^{3,4} and I now report that evolutionary history of the new millisecond pulsar is perhaps the simplest of them all. The rapid rotation of the pulsar, small mass function, circular orbit, and long binary period suggest that following the formation of a neutron star there had to be a phase of mass transfer from the low-mass secondary onto the neutron star primary. That mass transfer circularized the orbit and spun up the primary, and it required the secondary to be a red giant at that time. The present binary period of 120 days implies that the secondary has a mass of $0.3 M_\odot$. The observed mass function of $0.0027 M_\odot$ implies that the primary must be less massive than $3.1 M_\odot$, not an interesting mass limit for a neutron star. The secondary is probably a white dwarf at present. However, if the mass transfer terminated $< 10^6 \text{ yr}$ ago then the secondary may be still burning hydrogen at the rate of about $300 L_\odot$, and it may be optically detectable.

The original binary probably had a rather long orbital period and the initial masses had to be very different, such as $10 M_\odot$ and $1 M_\odot$. The more massive primary evolved all the way to supernova explosion and formation of a neutron star, while the secondary remained on the main sequence, a very standard evolutionary scenario⁴⁻⁶. It is expected that the orbit was highly eccentric following supernova explosion. After a long time, the secondary left the main sequence and became a red giant with degenerate helium core, hydrogen burning in a shell and an extended convective envelope. Gradual increase of radius led to mass transfer and the orbit became circular. The system looked like a model for a low mass X-ray binary described by Webbink *et al.*⁵. In fact the evolution of the binary was probably very close to that described with their Fig. 4. It is very difficult to predict what fraction of mass was transferred to the primary, and what fraction was lost from the binary during this evolution, but it is easy to describe the final stage of mass transfer and/or

mass loss, when the secondary was filling its Roche lobe for the last time. Following that, its mass remained constant, and certainly much smaller than the neutron star mass. Therefore, we may use a simple relationship between the secondary's mass, its Roche lobe radius and the binary period (equation (4) in ref. 7)

$$P(h) = 8.85(R_2/R_\odot)^{1.5}(M_\odot/M_2)^{0.5} \quad (1)$$

For a binary with a period of 120 days this may be written as

$$R_2 = 32 R_\odot (M_2/0.3 M_\odot)^{1/3} \quad (2)$$

at the last moment of mass transfer. At that time the secondary had the largest radius possible for a star of mass M_2 at the hydrogen shell burning stage. Subsequent hydrogen shell burning reduced the envelope mass and radius and the secondary evolved to become a white dwarf. A very similar evolutionary scenario was suggested⁸ long ago for AS Eri. The maximum stellar radii possible during the hydrogen shell burning phase for stellar remnants with total mass of 0.26, 0.35, and $0.44 M_\odot$ are 10, 55, and $135 R_\odot$, respectively⁹. The relation between the first two models may be approximated as

$$R_{2,\text{max}} = 23 R_\odot (M_2/0.3 M_\odot)^{5/7} \quad (3)$$

a very steep relation indeed. Combining equations (2) and (3) gives $0.32 M_\odot$ for the secondary at the end of mass transfer, suggesting that the secondary is probably now a white dwarf of $0.32 M_\odot$.

It is interesting that the secondary's mass was calculated using the value of a binary period only. Of course, the particular evolutionary scenario was considered because of a very small mass function, and the knowledge that the orbit is circular. It has been assumed that at least during the final stages of mass transfer from the secondary onto the more massive primary, the binary period and the secondary's Roche radius were increasing—a standard consequence of binary evolution—with a conservation of orbital angular momentum. As a result it was possible to adopt equation (3) for the secondary at the end of mass transfer. Note that a low mass of the secondary implies a rather long duration of the mass transfer, probably $> 10^7 \text{ yr}$ (Fig. 4 in ref. 3), plenty of time to spin up the neutron star.

As the inclination of the orbit is not known it is not possible to determine the mass of the pulsar. We can only write the standard expression for the mass function:

$$f(M) = M_2^3/(M_1 + M_2)^2 \sin^3 i = 0.0027 M_\odot, \\ \sin i = 0.44(M_1/M_\odot + 0.32)^{2/3} \quad (4)$$

A neutron star of $1.4 M_\odot$ or $2 M_\odot$ requires the inclination of 39° or 50° , respectively. The maximum mass allowed for the pulsar is $3.1 M_\odot$, corresponding to $i = 90^\circ$, not an interesting limit.

All the masses estimated here will need some adjustment when the observable binary parameters have been established with a higher accuracy. It is likely that the required changes will not be large, mainly because of a very steep maximum radius-mass relation for stars burning hydrogen in a shell (equation (3)). There is little hope of detecting a white dwarf companion at the expected distance of the millisecond pulsar. However, if the binary became detached only recently, then the secondary may be still burning hydrogen in the shell and it may be fairly bright. Interpolating between the published models of stars with masses $0.26 M_\odot$ and $0.35 M_\odot$ (ref. 9) we find that a star of $0.32 M_\odot$ needs $\sim 10^6 \text{ yr}$ to burn the residual hydrogen left in its envelope and to reduce its radius from $33 R_\odot$ while filling the Roche lobe down to about $0.05 R_\odot$, when it no longer can burn hydrogen. During those 10^6 yr the secondary should evolve horizontally across the H-R diagram at a constant luminosity of about $300 L_\odot$. If it is in that stage now, then it may be detectable in the optical part of the spectrum even at a large distance.

I thank Dr J. P. Ostriker for a stimulating discussion. This research was partly supported by NSF grant AST 80-22521. B.P. is a member of the Institute for Advanced Study, Princeton.

* Permanent address: N. Copernicus Astronomical Center, Polish Academy of Sciences, Warsaw, Poland.

Received 14 June; accepted 1 July 1983.

1. Boriakoff, V., Buccheri, R. & Fauci, F. *IAU Circ.* No. 3806 (1983); *Nature* **304**, 417–419 (1983).
2. Taylor, J. H. *IAU Symp.* No. 95, 361–369 (1981).
3. Blandford, R. D. & DeCampli, W. M. *IAU Symp.* No. 95, 371–377 (1981).
4. van den Heuvel, E. P. J. *IAU Symp.* No. 95, 379–394 (1981).
5. Webbink, R. F., Rappaport, S. & Savonije, G. J. *Astrophys J.* (in the press).
6. Trimble, V. *Nature* **303**, 137–142 (1983).
7. Paczynski, B. *Acta Astr.* **31**, 1–12 (1981).
8. Refsdal, S., Roth, M. L. & Weigert, A. *Astr. Astrophys.* **36**, 113–120 (1974).
9. Refsdal, S. & Weigert, A. *Astr. Astrophys.* **6**, 426–440 (1970).

Is the 6-ms binary pulsar the remnant of a bright galactic X-ray source?

G. J. Savonije

Astronomical Institute, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

It is argued here that the recently discovered 6-ms radio pulsar (ref. 1) with orbital period $P_b = 120 \pm 4$ days, $a \sin i = (0.92 \pm 0.08) \times 10^{12}$ cm and $e \approx 0$ is the spun up remnant of a formerly heavily accreting ($\dot{M} \approx 10^{-8} M_\odot \text{ yr}^{-1}$) X-ray source, belonging to a new type of low-mass X-ray binaries. This new type of X-ray binary consists of a critical lobe filling (lower) branch giant star of $\approx 1 M_\odot$ that spills mass to its neutron star companion. Systems of this kind were recently introduced² to explain the existence of the very bright ($L_x \approx 10^{38} \text{ erg s}^{-1}$) X-ray sources located in the bulge of M31 and in the bulge of our Galaxy. Such systems will end their lives as wide binary systems ($P_b \geq 100$ days), consisting of a $\approx 0.3 M_\odot$ helium white dwarf (the remnant core of the giant) in orbit about a rapidly spinning neutron star. For a suitable combination of magnetic field strength and accretion rate onto the neutron star, it will appear as a millisecond radio pulsar once the giant has lost its envelope and the heavy mass transfer ceases abruptly.

The observed small eccentricity ($e \approx 0$) of the wide binary orbit strongly suggests that the companion of the pulsar has been an evolved star with dimensions not much smaller than the orbital separation of the binary system. A critical lobe filling giant with a convective envelope is indeed thought to be very efficient in circularizing a wide binary orbit³. The envelope of this lobe filling giant decreases in mass by a combination of nuclear (hydrogen) shell burning at its base and tidal overflow at its surface. When the giant is less massive than its neutron star companion the mass transfer proceeds on the nuclear expansion time scale of the giant. It is known⁴ that the luminosity and radius of such a giant are virtually uniquely determined by the mass M_c of the degenerate helium core, that is, the structure of the giant's envelope is almost independent of its mass. Detailed calculations⁵ show this to be the case until the envelope mass has fallen to a few per cent of the total mass. It is easy to estimate the rate of mass loss $\dot{M}$ to the neutron star². After an initial decay the mass transfer rate appears to remain almost constant during most of the mass transfer phase. It appears² that the mean rate of mass transfer is linear proportional to the initial orbital period P_0 (at the beginning of the tidal overflow):

$$\langle \dot{M} \rangle \approx -5 \times 10^{-10} (P_0/\text{day}) M_\odot \text{ yr}^{-1} \quad (1)$$

The duration of the mass transfer (X-ray lifetime) is roughly given by:

$$\tau_x \approx 1 \times 10^9 (P_0/\text{day})^{-1} \text{ yr} \quad (2)$$

P_0 should be sufficiently long to permit the star to reach the base of the giant branch before tidal mass loss begins: $P_0 \geq 1$ day.

Almost the complete envelope of the giant (a small fraction is burned to helium and accreted by the degenerate core) will be transferred to the neutron star companion. The neutron star is therefore expected to spin up to roughly its equilibrium spin

period (at which accreting matter in a keplerian orbit near the boundary of the magnetosphere corotates with the spinning neutron star)⁶:

$$P_{eq} \approx 6(B_9)^{6/7} (R_{10})^{8/7} (M_p)^{-5/7} (\dot{M}_{17})^{-3/7} \text{ ms} \quad (3)$$

Here B_9 is the surface magnetic field strength of the neutron star in units of 10^9 G, R_{10} its radius in units of 10 km, M_p its mass in $M_\odot$ and $\dot{M}_{17}$ the accretion rate in units of 10^{17} g s^{-1} . The mass transfer ceases abruptly once the envelope is depleted below the critical envelope mass (a few per cent of the core mass). We are then left with a helium white dwarf of mass $0.2 \lesssim M/M_\odot \lesssim 0.4$ and a presumably rapidly spinning neutron star in a wide orbit.

The 6-ms pulsar seems to fit neatly into this evolutionary scenario. The mass of its companion follows as:

$$M_{wd} = a_p (P_b/0.116)^{-2/3} \dot{M}_i^{2/3} M_\odot \quad (4)$$

where a_p is the orbital distance from the pulsar to the centre of mass of the binary system in units of $R_\odot$, P_b the orbital period in days and $\dot{M}_i = \dot{M}_{wd} + \dot{M}_p$ the total binary mass in solar units. By adopting the canonical mass $M_p = 1.4 M_\odot$ for the pulsar and assuming a value for the orbital inclination of the binary, one can determine the mass of the companion M_{wd} , and the radius of the Roche lobe (R_c) around the latter star. Table 1 shows the results together with the calculated radius R_g of the giant (defined by its core mass M_{wd} , ref. 2) just before the collapse to a white dwarf configuration. By requiring $R_c \approx R_g$, there appears to be a unique solution for $\sin i \approx 0.6$ with $M_{wd} \approx 0.3$. (For a pulsar mass $M_p = 2.0 M_\odot$ these numbers become $\sin i \approx 0.75$ and $M_{wd} \approx 0.3$.) Thus it seems that the giant progenitor of the white dwarf had a core mass of $0.3 M_\odot$, implying² an initial orbital period $P_0 \approx 12$ days. We find with equations (1) and (2) that the progenitor binary was probably a bright low-mass X-ray binary ($\dot{M} \approx 6 \times 10^{-9} M_\odot \text{ yr}^{-1}$) during $\tau_x \approx 10^8$ yr. When the initial giant mass was $1 M_\odot$ it must have transferred $\approx 0.7 M_\odot$ to the neutron star. Only a small fraction of this mass (transferred at a rate below the Eddington limit) is required to spin the neutron star up to 6 ms (ref. 7), so that part of the transferred mass may have been ejected from the system by centrifugal forces near the magnetospheric boundary. Assuming that no appreciable spindown has occurred since the termination of the accretion phase (see discussion below), we adopt $P_{eq} \approx 6$ ms and $\dot{M} \approx 10^{-8} M_\odot \text{ yr}^{-1}$ at the end of the accretion (X-ray) phase. From equation (3) we derive a rough value of 3×10^9 G for the surface magnetic field strength at the birth of the radio pulsar phase. This is much smaller than the canonical value of 10^{12} G for young radio pulsars and suggests that substantial magnetic field decay has taken place since the formation of the pulsar. In view of the likely large age ($> 10^6$ yr) of the system, this is not surprising. That the field has not decayed even more may suggest that the neutron star was born relatively recently by accretion induced collapse of a white dwarf (see ref. 2 for possible evolutionary scenarios). From the very fact that the neutron star acts as a radio pulsar we

Table 1 System parameters as a function of orbital inclination i

$\sin i$	$M_{wd}(M_\odot)$	$a(R_\odot)$	$R_c(R_\odot)$	$R_g(R_\odot)$
1.0	0.171	119	26	1.8
0.9	0.191	119	27	3.2
0.8	0.217	120	28	6.1
0.7	0.255	120	29	14
0.6	0.303	121	31	33
0.5	0.370	124	34	92
0.4	0.482	126	37	379
0.3	0.689	130	41	—
0.2	1.19	141	51	—

The pulsar mass is adopted to be $1.4 M_\odot$. M_{wd} is the mass of the companion of the 6-ms pulsar, a the (total) orbital separation, R_c the radius of the Roche lobe around the companion¹² and R_g the radius of a giant ($Z = 0.02$) with core mass M_{wd} (ref. 2). There is a unique solution ($R_c \approx R_g$) for $\sin i \approx 0.6$.

conclude that the companion does not show substantial mass loss at present (white dwarf) and that the magnetic field apparently has not been aligned during the possibly long evolutionary history of the pulsar. If magnetic dipole radiation is the dominant energy loss mechanism of the pulsar at present, spindown occurs on a time scale⁹:

$$\tau_{sd} \approx 2 \cdot 10^7 \left(\frac{I_{45}}{B_{\perp}^2 R_{10}^6} \right) P^2 (\text{ms}) \text{ yr} \quad (5)$$

Here $B_{\perp}$ is the surface magnetic field component perpendicular to the pulsar's rotation axis in units of 10^9 G and I_{45} its moment of inertia in units of 10^{45} g cm². Substituting $P = 6.13$ ms and assuming $B_{\perp} = 3$, $I_{45} \approx 1$ and $R_{10} \approx 1$, yields $\tau_{sd} \approx 8 \times 10^7$ yr, consistent with the observed upper limit¹ of $\dot{P} < 6 \times 10^{-16}$. It is conceivable that the present pulsar magnetic field strength is even lower (we have to await a more accurate $\dot{P}$ measurement) and that the pulsar spin period P was even shorter at the end of the accretion phase. It cannot have been substantially shorter, because radiation of gravitational waves by non-radial oscillations of the rapidly spinning pulsar is expected to limit the spin period to ≥ 1.5 ms (ref. 8). It is, in fact, very unlikely that this limit was actually reached, as this requires (equation (3)) $B \approx 8 \times 10^8$ G. With such a low magnetic field strength spindown to 6 ms could not have occurred since the end of the accretion phase ($\tau_{sd} \approx 10^{10}$ yr).

A low magnetic field strength fits in nicely with the observed lack of X-ray pulsations from low mass X-ray binaries, because

the accreting plasma is then unlikely to be channelled to the magnetic poles and the accretion disk may extend almost down to the neutron star surface. It is difficult to estimate the galactic abundance of fast pulsar members of wide binary systems. The spindown time scale of the millisecond pulsar is rather long (equation (5)), so that the abundance of similar systems seems to be limited by the uncertain decay rate (or alignment time) of the pulsar magnetic field.

If one identifies the progenitor systems with bright bulge X-ray sources ($\tau_X \approx 10^8$ yr) the birth rate of millisecond pulsar systems must be very small (compare with ref. 13). There are only 30 X-ray sources with $L_X \geq 10^{37}$ erg s⁻¹ in our Galaxy¹⁰, ~20 of which are situated in the galactic bulge¹¹. Several of the bright non-bulge sources belong to the class of massive X-rays binaries, implying that there can only be very few low-mass giant X-ray binaries outside the galactic bulge (such as, Cyg X-2).

In this light it seems rather odd that the 6-ms pulsar happens to lie outside the galactic bulge and close to the galactic plane ($l'' = 65.8^\circ$, $b'' = 0.4^\circ$). This position suggests a rather massive parent system.

Although the observed 6-ms pulsar thus seems to fit in several respects neatly into the above sketched scenario of low-mass giant X-ray binary systems, there clearly remain questions, whose solution may shed new light on the nature of galactic X-ray sources.

I thank Dr Huib Henrichs for critical discussions.

7. Henrichs, H. F. & van den Heuvel, E. P. J. *Nature* **303**, 213–216 (1983).
8. Papaloizou, J. & Pringle, J. *Mon. Not. R. astr. Soc.* **184**, 501–508 (1978).
9. Pacini, F. *Nature* **216**, 567 (1967).
10. Clark, G., Doherty, R., Li, F., Jernigan, F. G. & van Paradijs, J. *Astrophys. J. Lett.* **221**, L37–L41 (1978).
11. Markert, T. H. et al. *Astrophys. J.* **218**, 801–814 (1977).
12. Eggleton, P. P. *Astrophys. J.* **268**, 368, 369 (1983).
13. Fabian, A. C., Pringle, J. E., Verbunt, F. & Wade, R. A. *Nature* **301**, 222–223 (1983).

Received 31 May; accepted 29 June 1983.

1. Boriakoff, V., Buccheri, R. & Fanci, F. *IAU Circ.* No. 3806 (1983).
2. Webbink, R. F., Rappaport, S. & Savonije, G. J. *Astrophys. J.* (in the press).
3. Zahn, J. P. *Astr. Astrophys.* **57**, 383–394 (1977).
4. Refsdal, S. & Weigert, A. *Astr. Astrophys.* **6**, 426–440 (1970).
5. Refsdal, S. & Weigert, A. *Astr. Astrophys.* **13**, 367–373 (1971).
6. Davidson, K. & Ostriker, J. P. *Astrophys. J.* **179**, 585–598 (1973).

X-ray emission and spin-up evolution of the 6.1-ms pulsar

David J. Helfand, Malvin A. Ruderman & Jacob Shaham*

Department of Physics and Columbia Astrophysics Laboratory, Columbia University, 538 West 120th Street, New York, New York 10027, USA

Neutron stars with rotation periods in the millisecond regime have been observed both as an isolated pulsar (PSR1937+214)¹ and as a pulsar with a low-mass binary companion (PSR1953+29)². We report here an upper bound to the X-ray luminosity of the latter of $L_X < 4 \times 10^{32}$ erg s⁻¹, and consider the consequences of this and other observational data for theoretical models of this pulsar's formation and spindown. The two fast pulsars have features in common which suggest a similar genesis (very low magnetic field and surface temperature, low galactic latitude) and we propose that both are members of a new class of radio pulsars³. We argue that their immediate ancestors were neutron stars in galactic bulge X-ray sources, which accrete from close, low-mass ($\leq 1 M_{\odot}$) binary companions and spin-up in the process. In this scenario, an isolated fast pulsar results from a tight binary where mass outflow from a (still) main sequence secondary is controlled by angular momentum loss from the binary motion. A wider binary, where Roche-lobe overflow is sustained by nuclear evolution of the secondary into a red giant phase, ultimately leaves behind an old spun-up pulsar with a light remnant companion. In both cases, the neutron-star progenitors would have been white dwarfs which collapsed under excess accreted mass^{4–7}. These neutron stars, then, are old ($> 10^7$ yr for binary pulsars, $> 10^9$ yr for isolated ones) which accounts for their anomalously low magnetic fields, and their consequent fast spins and low spindown rates.

In the course of a pointing at the known pulsar PSR1952+29, a field containing PSR1953+29 was imaged in the 0.25–3.5 keV X-ray band with the Imaging Proportional Counter onboard the Einstein Observatory⁸. The observation was carried out on 8 April 1979; 2,600 s of data were accumulated. No source is apparent at the position of the new pulsar. We derived an upper limit to the X-ray flux by summing the detected counts falling within a circle of radius 150 arc s and subtracting a local background determined from a concentric annulus of inner and outer radii 180 and 320 arc s, respectively: $S_X(0.25–3.5 \text{ keV}) < 0.005$ counts s⁻¹ (2σ). Assuming a Crab-like spectrum and correcting for absorption by a medium with a mean density of $n \sim 0.3$ cm⁻³ along the 3.5 kpc line-of-sight², this count-rate yields a flux limit of $f_X < 2.5 \times 10^{-13}$ erg cm⁻² s⁻¹; other spectral assumptions give flux values within 50%. The binary pulsar system X-ray luminosity, then, is $L_X(0.25–3.5 \text{ keV}) < 4 \times 10^{32}$ erg s⁻¹.

Using the methods outlined in ref. 9 we can translate the observed X-ray flux limit into a limit on the temperature of the surface of PSR1953+29; for the same line-of-sight interstellar density ($n \sim 0.3$ cm⁻³) and a neutron star radius of 15 km, $T < 1 \times 10^6$ K. When coupled with recent neutron star cooling calculations, this provides a weak lower limit on the age of the pulsar $t \geq 3,000$ yr.

We discuss below several less trivial consequences which also depend, directly or indirectly, on the above observational bound on L_X : (1) the neutron star in the binary is really very old ($> 10^7$ yr); (2) $\dot{P}$, the spin-down rate, is $\leq 10^{-17}$ ss⁻¹ (the reported upper bound² is $\sim 6 \times 10^{-16}$ ss⁻¹); (3) the neutron star was not made in a conventional supernova explosion involving large mass expulsion; and (4) the neutron star was probably spun up by accretion during the red giant phase of its companion and is another member of the class of 'resurrected pulsars'¹⁰.

PSR1953+29 has a companion of $\sim 0.3 M_{\odot}$, revolving in a low-eccentricity orbit with a radius of $\sim 10^{13}$ cm (ref. 2). There are two possible scenarios for neutron star formation in a binary system: (1) core collapse in a massive star ($> 4 M_{\odot}$) leading to

* Permanent address: Racah Institute of Physics, Hebrew University, Jerusalem, Israel.

a supernova (SN) explosion, and (2) collapse of a white dwarf pushed over the Chandrasekhar limit by accretion from its companion. We consider below which of these is likely to produce a system with the parameters of PSR1953+29 and explore the limits these parameters impose on the detailed history of this object.

In scenario (1), the neutron star, of mass m_{NS} , was formed when a star of mass m'_1 , orbiting a star of mass m'_2 in a nearly circular orbit, collapsed in a conventional supernova (SN) explosion. Ignoring any induced changes of m'_2 , the sudden mass ejection of $\Delta m_1 (= m'_1 - m_{\text{NS}})$ is related to the resulting orbital eccentricity e (< 1) through

$$\Delta m_1 = m'_T \frac{e}{1+e} < \frac{1}{2} m'_T \quad (1)$$

where $m'_T = m'_1 + m'_2$. For conventional supernovae $\Delta m_1 > m_{\text{NS}}$, so that a companion of $\sim 0.3 M_\odot$ would have necessarily become unbound, even if its orbit was inside of the pre-SN giant's envelope. Only a much heavier companion would have survived the explosion, and it would probably be left moving in an eccentric orbit. Thus, the age of the present (neutron star plus low-mass star) binary is the larger of the evolution time of the companion to reach $\sim 0.3 M_\odot$ or the circularization time of the orbit. At the current separation, mass loss from the secondary could have taken place only through a red giant phase, during which the mass lost was partly accreted onto the neutron star on a time scale $> 10^7$ yr (ref. 11). To this, one must add the secondary's main sequence evolution time remaining at the time of the SN. Circularization times of such systems are also $> 10^7$ yr (refs 12, 13). Thus, the SN must have occurred at least 10^7 yr ago; scenario (1) requires that PSR1953+29 is old.

This picture suffers from two severe difficulties as an explanation for the origin of pulsars such as PSR1953+29. The expected recoil velocity of the centre of mass of the remaining binary, following the SN, is

$$v_{\text{cm}} = ev'_1 \quad (2)$$

where v'_1 is the pre-SN orbital velocity of m'_1 . If the giant secondary lost its envelope in a steady fashion, its Roche lobe should have been constantly expanding; hence the binary separation at the time of the SN should have been $\sim 10^{12}$ cm. At present, $v_2 \sim 60 \text{ km s}^{-1}$ so that one might expect $v_{\text{cm}} \sim 100 \text{ km s}^{-1}$. This, of course, is comparable to the typical observed velocities of normal pulsars, which are usually presumed to be runaway SN remnants from just such binaries which do not remain bound. In the $> 10^7$ yr that has passed since the SN in our first scenario, then, the still-bound binary would have thus travelled $> 3 \times 10^{21} \text{ cm} \sim 1 \text{ kpc}$ from its birthplace (near the galactic plane) to a galactic latitude, at 3.5 kpc, of $b \sim 20^\circ$. The measured value of b for PSR1953+29 is 0.4° . The three other known binary pulsars and the original millisecond pulsar are also all $< 300 \text{ pc}$ from the plane with a mean $|z| = 130 \text{ pc}$.

Another argument against this model concerns a distinctive characteristic of the four radio pulsars that are members of binary systems: all have extremely weak magnetic fields. In scenario (1), the supernova explosion is not triggered by the state of the companion star. We would expect, then, that at birth, such pulsars would be orbited by a main sequence star that neither fills its Roche lobe, nor possesses a strong enough stellar wind to suppress the pulsar's radio emission. Thus, the more commonly observed low-mass binary should contain a conventional young, high-field pulsar (similar to Vela) rather than one with an anomalously high rotation rate and an anomalously low magnetic field. If the binary characteristic of the pulsar progenitor's environment is to influence the pulsar's properties, the companion must affect the progenitor's evolution; in any type (1) scenario, it does not do so.

We shall, therefore, emphasize here the (more likely) second scenario for the formation of PSR1953+29.

In scenario (2), the neutron star's progenitor was a white dwarf near its Chandrasekhar limit. It accreted mass from its

binary companion's Roche-lobe overflow once the companion left the main sequence and began to expand as a red giant. When the dwarf's mass m'_1 reached $1.3 M_\odot$, the Fermi energy of its core electrons grew large enough to allow electron capture by core nuclei (O, Ne, Mg) and thus trigger a contraction to a neutron star. This contraction might be quite slow if the dwarf has accreted enough mass from a keplerian accretion disk fed by its companion so that centrifugal force brakes the collapse, or it may be a sudden implosion. In the former case, any escaping mass would be ejected gradually (on a time scale much longer than a binary period) so that almost no binary recoil is expected; in the latter case, the total ejected mass has been estimated to be less than a few tenths of a solar mass⁹. Thus, quite generally, the binary recoil velocity from the formation of the neutron star is $< 10^{-1} v'_1 \sim 10 \text{ km s}^{-1}$, comparable to the random z -motion of Population I stars at formation. The binary would then oscillate about the galactic plane with a maximum excursion $z_m = (v'_1)^2 / 2k$ where $k \sim 5 \times 10^{-9} \text{ cm s}^{-2}$ in the solar neighbourhood. Thus, $z_m \sim 100 \text{ pc}$ and $|b|_{\text{max}} \sim 0.6^\circ$ for PSR1953+29, not inconsistent with its observed galactic latitude or with that of the other binaries and PSR1937+214. However, sudden implosion will introduce an eccentricity which may not be easy to get rid of^{12,13}.

If there is either negligible or sufficiently slow mass ejection when the neutron star forms, the orbital parameters of the system and the atmosphere of the giant companion will not change significantly during the process. Following the collapse, the pre-SN accretion of the white dwarf will continue onto the newly formed neutron star until only the degenerate core of the companion remains. Only then will the neutron star be observable as a radio pulsar. If, on the other hand there is a sudden mass ejection of a few tenths of a solar mass in the white dwarf-neutron star transition, there will be a sudden expansion of the companion's Roche lobe. Rapid mass transfer will then cease until, on a nuclear-burning time scale, the giant phase companion grows large enough to refill its expanded Roche lobe.

Although, in principle, the newly formed neutron star might be observable as a radio pulsar during this detached phase, the present bound on the X-ray luminosity of PSR1953+29 argues strongly against the possibility that we are observing it in that interval. The giant star, even though it no longer fills its newly expanded Roche lobe, would still be losing mass in a stellar wind at a rate $\dot{M}_w$ given approximately by¹⁴

$$\dot{M}_w \sim 4 \times 10^{-14} (R/R_\odot)(L/L_\odot)(M/M_\odot)^{-1} M_\odot \text{ yr}^{-1} \quad (3)$$

of which a fraction

$$f \sim \left[\frac{Gm_1}{v^2 a} \right]^2 \quad (4)$$

would be captured by the neutron star, where v is the quadrature sum of the wind velocity v_w and the companion's velocity v_2 . For $v_w = v_2$, $f \sim 0.25$ and the neutron star X-ray luminosity, given by $L_x \sim 0.1 \dot{M}_w c^2$, is $L_x \sim 10^{20} |\dot{m}_2| \text{ erg s}^{-1}$. This value for L_x equals that of the observational X-ray bound for $|\dot{m}_2| \sim 2 \times 10^{13} \text{ g s}^{-1} \sim 4 \times 10^{-13} M_\odot \text{ yr}^{-1}$, far less than $\dot{M}_w$ from equation (3) for any late-type giant.

We shall, therefore, turn away from the suggestion that PSR1953+29 is young and consider the remaining and surviving hypothesis, that we have now observed the pulsar when it is well over 10^7 yr old, that is, after the red giant phase of its companion has ceased. At this age, theory and observation indicate that currents in the neutron star's crust should have died out. While currents in the liquid core remain, magnetic field rearrangements there are no longer restrained by crustal conductivity leading to a very much reduced external magnetic field.

During the long ($\geq 3 \times 10^8$ yr) giant Roche-lobe overflow phase, the neutron star is accreting at $\sim 10^{-8} M_\odot \text{ yr}^{-1}$, and is observable as one of the bright galactic bulge X-ray sources (most of which also lie very close to the galactic plane). The accreted angular momentum brings the neutron stellar spin to a steady state with its surrounding keplerian accretion disk;

only $\sim 0.04 M_{\odot}$ of accreted matter will suffice. When accretion ceases and the neutron star is observable as a radio pulsar, it has been shown³ that the initial spindown rate is given by

$$\dot{P}_0 \sim 10^{-19} \text{ ss}^{-1} (10^3 P_0)^{4/3} (\dot{m}_1 / 10^{-9}) M_{\odot} \text{ yr}^{-1} \quad (5)$$

where P_0 is the initial period in seconds. In equation (5) we have assumed $m_{\text{NS}} \sim 1.4 M_{\odot}$ and $I_{\text{NS}} = 10^{45} \text{ g cm}^2$. If $\dot{m}_1$ is limited to the Eddington rate of $10^{-8} M_{\odot} \text{ yr}^{-1}$, the upper bound for PSR1953+29 is

$$\dot{P}_0 \leq 10^{-17} \text{ ss}^{-1} \quad (6)$$

about two orders of magnitude less than the current observational upper limit. Since the pulsar may have already lived several $P_0/\dot{P}_0$ time units, we may expect $\dot{P} < \dot{P}_0$.

The lack of observed X-ray emission from the pulsar and its neighbourhood supports the prediction that $\dot{P}$ is indeed bounded by equation (6). A half-dozen pulsars with spin-down energy loss rates

$$\dot{E} = I \Omega \dot{\Omega} = (2\pi)^2 I \dot{P} P^{-3} \geq 10^{34} \text{ erg s}^{-1} \quad (7)$$

were observed with Einstein and found to be surrounded by diffuse X-ray nebulae^{15,16}. This is presumably due to the acceleration of relativistic particle beams into the surrounding interstellar medium. In fact, the only pulsar with $\dot{E} > 10^{34} \text{ erg s}^{-1}$ not embedded in such a nebula is the very low-field 'millisecond pulsar' PSR1937+24, for which $\dot{E} \leq 10^{36} \text{ erg s}^{-1}$. The spin-down energy loss rate of PSR1953+29, which also has no observed nebula, is

$$\dot{E} \sim \left(\frac{\dot{P}}{10^{-15} \text{ ss}^{-1}} \right) \times 10^{38} \text{ erg s}^{-1} \quad (8)$$

For this to be no greater than the upper bound to that of PSR1937+24, $\dot{P} \leq 10^{-17} \text{ ss}^{-1}$, about two orders of magnitude less than that of its present upper bound.

We note finally, that a $\dot{P} \sim 6 \times 10^{-16} \text{ ss}^{-1}$ for a pulsar with $P \sim 6 \times 10^{-3} \text{ s}$ implies a pulsar lifetime at that period of $P/\dot{P} \sim 3 \times 10^5 \text{ yr}$. If one such pulsar has now been observed, one would expect to have observed ~ 100 descendants of such pulsars with periods $P \leq 6 \times 10^{-2} \text{ s}$ and $\dot{P} \geq 6 \times 10^{-17} \text{ ss}^{-1}$, for which pulsar turnoff has not yet been reached (see Fig. 1 of ref. 3). By contrast, a $\dot{P} \leq 10^{-17}$ for PSR1953+29 would imply that all of these many $6 \times 10^{-2} \text{ s}$ descendants have $\dot{P} \leq 10^{-18} \text{ ss}^{-1}$ and hence have already turned off as radio pulsars.

A significant lower bound to the $\dot{P}$ of PSR1953+29 would follow if the Cos B γ -ray source 2CG065+00 is indeed the pulsar at a distance of 3.5 kpc. Its γ -ray luminosity would then be comparable with that of the Crab pulsar and, to supply this luminosity, one would need $\dot{P} \geq 10^{-17}$. When combined with equation (6) or the above additional arguments giving upper bounds to $\dot{P}$, we are led to predict that either $\dot{P} \sim 10^{-17}$ or the existence of PSR1953+29 in the 2CG065+00 error box (which actually led to the search for the pulsar) is coincidental.

The present model for the formation and evolution of the binary pulsar PSR1953+29 thus begins with a cataclysmic variable turning, through the quiet collapse of the white dwarf, into a bright galactic bulge X-ray source. The neutron star then accretes near its Eddington limit and gets spun up to $\sim 6 \text{ ms}$ because its surface field stabilizes at $\sim 10^9 \text{ G}$ (ref. 3). The present companion started out as a solar-type main sequence star with an orbital period of several weeks.

The proposed evolutionary scenario³ for PSR1937+214, which left it with no observable companion, differed in that the initial mass loss X-ray binary was lighter and tighter. Its evolution was determined by the rate at which the binary contracted from angular momentum loss (magnetic torque from winds, gravitational radiation) so that the secondary lost most of its mass long before its core nuclear burning could convert it to a giant. It has been argued elsewhere¹⁷ that when the companion mass became sufficiently small, stable mass transfer was no longer possible and the (now degenerate) companion was tidally disrupted. In such cases, with initially contracting Roche lobes, mass transfer to the neutron star during the X-ray binary phase

will be, for most of its lifetime, less than in the case of the giant companion overflowing an expanding Roche lobe, but would last longer ($\leq 10^9 \text{ yr}$). The spun-up pulsar would then have a smaller $\dot{m}_1 \leq 10^{-9} M_{\odot} \text{ yr}^{-1}$ for equation (5), as indeed seems to be the case for PSR1937+214, a very close cousin of PSR1953+29 in this new class of old radio pulsars.

The authors acknowledge the support under NSF PHY80-23721, AFOSR82-0014 and NAS8-30753. This is Columbia Astrophysics Laboratory contribution no. 246.

Received 6 June; accepted 1 July 1983.

1. Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. *Nature* **300**, 615–618 (1982).
2. Boriakoff, V., Buccheri, R. & Fauci, F. *IAU Circ.* No. 3806 (1983).
3. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728–730 (1982).
4. van den Heuvel, E. P. J. *Ann. N.Y. Acad. Sci.* **302**, 14 (1977).
5. Miyajii, S., Nomoto, K., Yokoi, K. & Sugimoto, D. *Publ. astr. Soc. Jap.* **32**, 303 (1980).
6. Sugimoto, D. & Nomoto, K., *Space Sci. Rev.* **25**, 155 (1980).
7. Nomoto, K. in *Proc. Workshop on Type I supernovae*, Austin (1980).
8. Giacconi, R. *et al. Astrophys. J.* **230**, 540–550 (1979).
9. Helfand, D. J., Chanan, G. A. & Novick, R. *Nature* **283**, 337–343 (1980).
10. Ruderman, M. A. & Shaham, J. *Comments Astrophys.* (in the press).
11. Webbink, R. F., Rapport, S. & Savonije, G. J. *Astrophys. J.* (in the press).
12. Press, W. H., Wiita, P. J. & Smarr, L. L. *Astrophys. J. Lett.* **202**, L135–L137 (1975).
13. Lecar, M., Wheeler, J. C. & McKee, C. F. *Astrophys. J.* **205**, 556–562 (1976).
14. Reimers, D. *Mem. Soc. R. Sci. Liege* 6th ser **8**, 369–374 (1983).
15. Cheng, A. & Helfand, D. J. *Astrophys. J.* (in the press).
16. Helfand, D. J. *Supernova Remnants and Their X-ray Emission*, IAU Symp. No. 101 (in the press).
17. Ruderman, M. A. & Shaham, J. Preprint (Columbia University, 1983).

Fate of very low-mass secondaries in accreting binaries and the 1.5-ms pulsar

Malvin A. Ruderman & Jacob Shaham*

Department of Physics and Columbia Astrophysics Laboratory, Columbia University, New York, New York 10027, USA

Families of old tight low-mass binaries in which mass transfer to a heavier collapsed primary from a lighter secondary is an essential evolutionary feature include the galactic bulge X-ray sources, in which the primary is a neutron star. Canonical theories for the evolution of such binaries assume that the mass transfer and secondary evolution are always stable. We argue here that the stability postulate may fail if the secondary mass falls below $\sim 2 \times 10^{-2} M_{\odot}$. The secondary may then be almost entirely tidally disrupted, leaving behind, at most, a small remnant very much lighter than the Earth in an orbit with a period of several hours. In binary accretion scenarios for the origin of the millisecond pulsar PSR1937+214¹, this could explain the absence of any detected binary companion.

In canonical theories² the (orbital) angular momentum L is generally defined as that residing entirely in the motion of the primary and secondary about the system centre of mass. The key adjustable evolutionary parameter (h) contains the ratio of the orbital angular momentum loss rate from the two-body system ($\dot{L}$) to the mass loss rate from the secondary ($\dot{m}_2$): $h \equiv (\dot{L}/L)/(\dot{m}_2/\dot{m}_2)$ where $\dot{m}$ is the reduced mass $m_1 m_2 / (m_1 + m_2)^{1/2}$. Angular momentum expelled by mass flow from the binary system, or stored in an accretion disk within the system, or transferred (permanently) to stellar spin are not included in L but losses to these sinks do contribute to $\dot{L}$ as does binary angular momentum loss carried away by gravitational radiation. For $m_2 \ll m_1$, the case for low-mass accreting binaries near the end of their evolution, the orbital angular momentum of the binary is essentially (to order m_2/m_1) all in the secondary's motion. In these conditions, the sign of $h-1$ determines whether the binary separation (a) is increasing or decreasing:

$$\frac{\dot{a}}{a} \approx 2(h-1) \frac{\dot{m}_2}{m_2} \quad (1)$$

* Permanent address: Racah Institute of Physics, Hebrew University, Jerusalem, Israel.

The distance between the lighter secondary's centre of mass and its inner Lagrange point—a measure of the radius of its Roche lobe—is³

$$R_L = \eta a \left(\frac{\tilde{m}}{3m_1} \right)^{1/3} \quad (2)$$

where η is of order unity and equals one when almost all of m_2 is distributed spherically symmetrically. When the secondary is in a contracting low-mass binary, R_L decreases with time because both a and m_2 grow smaller. It can increase only for an expanding binary. Stable evolutionary scenarios are such that the secondary never grows larger than its Roche lobe and exactly fills it during any accretion phase. For a secondary with a mass (m_2)–radius (R_2) relation $R_2 = f^{-1} m_2^n$ the stability postulate is canonically taken to mean $R_2 = R_L$. When the secondary is very light, $m_1 \gg m_2 \sim \tilde{m}$, with density ρ , this implies

$$h \sim 1 - \frac{m_2}{6\rho} \frac{d\rho}{dm_2} \sim \frac{5-3n}{6} \quad (3)$$

where the density derivative is taken along the secondary's evolutionary path and f is assumed constant. Then equation (1) becomes

$$\frac{\dot{a}}{a} \sim -(n+1/3) \frac{\dot{m}_2}{m_2} \quad (4)$$

A light Roche lobe-filling main sequence secondary has an initial index $n \sim -0.8$, $h \sim 1.2$ and hence $\dot{a} < 0$. As evolution and mass loss proceed, the index will increase to $n = +1/3$, that for a nonrelativistic, cool degenerate dwarf² (and also that for a light, fully convective star). As the growing n passes through $n = -1/3$, $h \sim 1$ and $\dot{a}$ reverses from contraction to expansion. At this point a minimum binary period is achieved, and agreement between computed and observed minimum periods for such binary systems has encouraged support for the stable evolutionary model⁴. Afterwards, as $n \rightarrow 1/3$ and $h \rightarrow 2/3$, the binary begins to expand. The key question that we are raising here is—does it expand fast enough to accommodate the expanding R_2 when $m_2 \ll m_1$? Because that part of h which comes from gravitational radiation varies inversely with $\dot{m}_2$, the Roche lobe overflow rate can, in principle, adjust itself to achieve any needed value of $h > h_M$ where h_M includes only mass flow effects. Thus, in the case of constant f , if $h_M < 2/3$, canonical stable binary evolutionary scenario models are consistent along their entire evolutionary path. But if, say, $h_M \geq 2/3$, they are not. Rather, catastrophic instability and tidal disruption of the secondary might even be expected at some point in the late expanding binary phase because, by equation (3), one then needs $h \sim 2/3$ to maintain state equilibrium.

Therefore, stable low-mass binary evolution models hypothesize, based on a large number of simulations and observations, a sufficiently small h_M . Most common, perhaps, are fully or nearly 'conservative' models⁵ with an assumed $h_M \sim 0$. However, there are not yet available realistic models which can follow in quantitative detail the formation and evolution of accretion disks around the primary or the flow of mass ejected from the binary. The 'conservative' view deposits angular momentum lost to and by the accretion disk right back into the $m_1 m_2$ orbital motion via either torques or direct mass deposit. But some simple considerations below do suggest that $h_M \leq 2/3$ may not be achievable after the diminishing m_2 has fallen below $\sim 2 \times 10^{-2} M_\odot$.

The mass loss $\dot{m}_2$ occurs through the vicinity of the inner lagrangian point, a distance from the secondary's centre of angular momentum which approaches R_L of equation (2) when $m_2/m_1 \ll 1$. If the lost mass had no further effect on the orbital angular momentum of the binary then, for sufficiently small m_2/m_1 ,

$$h_M = \left(1 - \frac{R_L}{a} \right)^2 \sim 1 - 2 \left(\frac{m_2}{3m_1} \right)^{1/3} \quad (5)$$

This, of course, is not upper bounded by $2/3$. The question, therefore, is how will the h_M of equation (5) be changed (by

Δh_M) because of the subsequent interaction of the remnant m_2 with the gas stream formed from its ejecta, the ring formed by the gas stream and the accretion disk fed by the ring.

In steady-state accretion with very small m_2/m_1 , the m_2 –gas interaction may be separated into two contributions: first, the long-range gravitational interaction of m_2 with the ring and the bulk of the accretion disk; second, the shorter range (and mainly collisional) interaction of m_2 with any local keplerian disk gas whose orbits extend beyond the inner lagrangian point. Part of the first is essentially a tidal dissipation effect on the viscous keplerian disk, whose angular speed increases inwards, similar to tidal effects of the Moon on the more rapidly spinning Earth. The resulting torque on m_2 is of order $(m_2/m_1)G(m_d m_2/a)\theta$ where m_d is the disk mass and θ is the angle between the tidal bulge principal axes and the $m_1 m_2$ separation. The angle $\theta \sim \eta a^{-2} \Omega^{-1}$ for disk viscosity η and Ω , the orbital annular frequency of m_2 , but the product $\eta m_d a^{-2} \sim \dot{m}_2 \ll |\dot{m}_2|$. Then we have for the disk tidal contribution to equation (5) $\Delta h_{M,d} \sim (m_2/m_1)^2$ which is not important in our low m_2/m_1 limit.

The net gravitational torque between m_2 and the jet-ring system feeding the disk is not greater than $Gm_2 m_j a^{-1}$ where the asymmetric part of the 'jet' mass $m_j \leq |\dot{m}_2| a (R_L \Omega)^{-1}$ since $R_L \Omega$ is a measure of the relative velocity between the inner lagrangian point and the thin ring which carries away the ejected overflow stream which comes from it. Then $\Delta h_{M,j} \sim (m_2/3m_1)^{2/3}$ which is also of higher order in the assumed small parameter $(m_2/3m_1)^{1/3}$ of equation (5). Finally, for $m_2 \ll m_1$, m_2 may actually lie within the disk, part of whose matter always drifts outwards. Radiation and thermal pressure on the accretion disk may cause it to rotate slightly more slowly around m_1 than does m_2 . But the consequences of any disk– m_2 interaction resulting from that could only be an increase in h relative to that computed when ignoring this effect. Furthermore, collisions between m_2 and gas streams which may run into it can be shown to give contributions to h_M of the order of $(m_2/3m_1)^{2/3}$ which are again small relative to larger $(m_2/3m_1)^{1/3}$ terms in our low-mass limit. Thus none of the above mechanisms for interaction of the mass overflow in binaries with very light secondaries will achieve the $h_M < 2/3$ needed for a stable expanding binary phase model unless

$$2(m_2/3m_1)^{1/3} > 1/3 \quad (6)$$

For a $1.4 M_\odot$ primary white dwarf or neutron star, this criterion gives $m_2 > 2 \times 10^{-2} M_\odot$. In typical stable mass transfer scenarios for gravitational–radiation-driven mass transfer², such low m_2 can be reached within the Hubble time, while accretion still takes place, and tidal disruption may then ensue.

A tidal disruption of an $n = +1/3$ degenerate dwarf could proceed on a hydrodynamic time scale (~ 1 h) if the secondary's volume does indeed grow more rapidly than its Roche lobe in the expanding binary. This process could stop when solid-state forces become more important than gravitational ones in the remnant, so that $d\rho/dm_2 \rightarrow 0$ in equation (3) as $h \rightarrow 1$. This possible light planet remnant would fragment into a Saturn-like ring of 'rocks' if subject to net tidal tension. Therefore, the binary separation a would still have to satisfy $m_1 a^{-3} \sim 3\rho/4\pi$ or $a \sim 2 \times 10^{11}$ cm with a binary period of a few hours. If this is the case, we note that, for relatively low-pressure cool matter with an equation of state $d\rho/d\rho = Y/\rho_0$, compressibility modulus Y , zero pressure density ρ_0 and pressure $p = 2\pi/3 G \rho_0^2 R_2^2$, stable mass loss can be restored when $h_M = 1 - 2(m_2/3m_1)^{1/3} \leq 1 - m_2/6\rho d\rho/dm_2$. This gives an upper bound for the remnant mass $m_2 \leq 10^3 Y^3 (m_1 G^3 \rho_0^4)^{-1}$. With $Y/\rho_0 \sim 10^{10}$ cm² s⁻² (the squared zero-pressure sound speed), $m_1 \sim 1.4 M_\odot$ and zero-pressure matter density $\rho_0 \sim 3$ g cm⁻³, $m_2 \leq 10^{21}$ g. Such a light 'planet' or large asteroid with a several hour period, remaining after the catastrophic tidal disruption of the secondary, would be an undetectable residual companion to PSR1937+214 but if $h_M \geq 1$ as $m_2 \rightarrow 0$, no planetary residual of any sort would be left. We also note that as long as the low-mass secondary remnant is not light enough to satisfy $m_2 \leq 2 \times 10^{-2} M_\odot$, as is the case for the measured cataclysmic

variables and galactic-bulge X-ray sources, then the canonical scenario, leading to an expanded undisrupted binary, need not be at all inconsistent. Only very light companions are relevant to our discussion here.

Finally, there are additional possibilities for complete disruption of a galactic-bulge X-ray binary or tidal dissipation of its secondary even before the instability stage discussed above is reached. The high abundance of X-ray low-mass accreting binaries in globular clusters has been attributed to the higher probability for forming tight binaries involving a neutron star in that environment. But while still in the cluster, the accreting neutron star also has a greatly increased probability of losing the low-mass secondary which feeds it⁶. The close passage of another star can disrupt the binary or greatly accelerate the tidal breakup of the secondary in the expanding phase where, as we have seen above, it is always expected to be at least close to such an instability.

Thus the absence of any observed residual companion around a putative accretion spun-up PSR1937+214, or around various other isolated low magnetic field rapid pulsars which have been suggested as having a similar origin, does not appear to embarrass the proposed spin-up models. However, the above instability arguments would not apply for the heavier companion of the 6-ms binary pulsar^{7,8}.

We thank H. Henrichs, E.P.J. van den Heuvel, J. Grindlay, I. Iben, P. C. Joss, J. H. Krolik, B. Paczynski, K. H. Prendergast, R. F. Webbink, and A. V. Tutukov for many helpful discussions. This work was supported by NASA under contract NAS 8-30753 and by the NSF under grant PHY 80-23721. This is Columbia Astrophysics Laboratory contribution no. 243.

Received 26 April; accepted 5 July 1983.

1. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728–730 (1982).
2. See, e.g., Rappaport, S., Joss, P. C. & Webbink, R. F. *Astrophys. J.* **254**, 616 (1982).
3. Kopal, Z. *Close Binary Systems* (Chapman and Hall, London, 1959).
4. Henrichs, H. & van den Heuvel, E. P. J. *Nature* **303**, 213–216 (1983).
5. Faulkner, J. *Proc. IAU Symp.* No. 73 (1976).
6. Krolik, J. H., Meiksin, A. & Joss, P. C. Preprint (MIT, 1983).
7. Boriakoff, V., Bucccheri, R. & Fauci, F. *Nature* **304**, 417–419 (1983).
8. Helfand, D. J., Ruderman, M. A. & Shaham, J. *Nature* **304**, 423–425 (1983).

Aqueous-phase source of formic acid in clouds

W. L. Chameides & D. D. Davis

School of Geophysical Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

The concern over the environmental effects of acidic rain has increased the interest in understanding the processes which control the acidity of cloud and rainwater^{1–6}. In regions affected by anthropogenic emissions, H₂SO₄ and HNO₃ are most often responsible for lowering the pH of rain below 5.6, the value water attains in equilibrium with atmospheric CO₂ (ref. 7). In more remote regions, however, formic acid (HCOOH), and to a lesser extent, acetic acid (CH₃COOH), have also been identified as major acidic components of rain⁸. These two organic acids have also been observed in the gas phase in the southwestern US⁹. While the sources of HNO₃ and H₂SO₄ are at least qualitatively understood^{5,6,10–12}, the sources of organic acids remain largely unknown. We have investigated the coupled gas- and aqueous-phase cloud chemistry of HCOOH and report here that during the daylight hours, aqueous-phase OH radical reactions can both produce and destroy HCOOH in cloud droplets and may, in fact, control the HCOOH levels in rain. Similar mechanisms may also exist for acetic and other organic acids.

During the daytime, gaseous OH and HO₂ radicals are photochemically produced in the atmosphere^{10–12}. Within a cloud, these radicals can be rapidly scavenged by cloud droplets, thereby resulting in a source of free radicals to cloudwater¹³. Once in the aqueous phase, OH and HO₂ initiate a complex

series of reactions. As the details of this chemistry have been presented previously¹³, only the central features of the OH/HO₂ aqueous-phase system are described below (see Fig. 1).

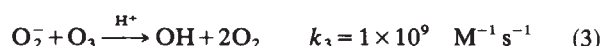
On dissolution, the HO₂ radical rapidly equilibrates with O₂⁻ (ref. 14) so that



These two species then react to form hydrogen peroxide by reactions such as¹⁴



or are converted to OH by reacting with dissolved O₃ through¹⁵

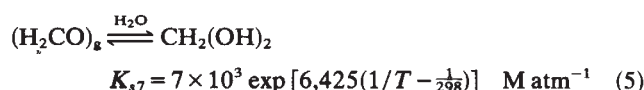


The dissolved OH species is extremely reactive and interacts with a wide variety of species including HSO₃⁻, the dissolved form of SO₂, through¹⁶



Not only does reaction (4) represent a major sink for OH in cloud droplets, but as SO₃⁻ is rapidly converted to SO₄²⁻, this reaction can also be an important pathway by which S(IV) is oxidized in a cloud to S(VI) (ref. 13).

In addition to reacting with HSO₃⁻, the OH radicals can take part in a reaction sequence which we refer to here as the aqueous-phase formaldehyde/formic acid oxidation scheme. This scheme is initiated by the dissolution of gaseous formaldehyde, (H₂CO)_g, into cloud droplets to form hydrated formaldehyde; such that^{17,18}



where $K_{s,7}$ represents the solubility or Henry's law constant for formaldehyde. It has been suggested that dissolved formaldehyde is photolysed in cloudwater to form CO (ref. 19).

Table 1 Conditions adopted for cloud model calculations

General parameters

Solar zenith angle = 30°
 $T = 295 \text{ K}$
 Relative humidity = 100%
 ϵ = cloud transmissivity = 0.5
 α = accommodation coefficient = 0.01
 N = concentration of droplets = 120 cm^{-3}
 a = radius of drops = $10 \mu\text{m}$
 W_L = liquid water content = 0.5 g m^{-3}

Gas-phase initial concentrations*

$\text{O}_3 = 3 \times 10^{-8} \text{ v/v}$
 $\text{CH}_4 = 1.65 \times 10^{-6} \text{ v/v}$
 $\text{NO}_x = \text{NO} + \text{NO}_2 = 1.5 \times 10^{-11} \text{ v/v}$
 $\text{HNO}_3 = 2.0 \times 10^{-11} \text{ v/v}$
 $\text{H}_2\text{O}_2 = 1.1 \times 10^{-9} \text{ v/v}^\dagger$
 $\text{SO}_2 = 5.5 \times 10^{-11} \text{ v/v}$
 $\text{NH}_3 = 5 \times 10^{-11} \text{ v/v}$
 $\text{H}_2\text{CO} = 3.8 \times 10^{-10} \text{ v/v}^\dagger$
 $\text{CO}_2 = 3.3 \times 10^{-4} \text{ v/v}$
 $\text{HCOOH} = 0.0 \text{ v/v}^\ddagger$

Aqueous-phase initial conditions

$[\text{H}^+] = 2 \times 10^{-6} \text{ M}$
 $[\text{HCO}_3^-] = 2 \times 10^{-6} \text{ M}$
 $[\text{Cl}^-] = 1 \times 10^{-4} \text{ M}$
 $[\text{Na}^+] = 1 \times 10^{-4} \text{ M}$

* Initial gas-phase concentrations correspond to those that might typically be encountered in the remote, marine boundary layer in cloud-free conditions and a relative humidity of 80%.

† These values were determined from photochemical modelling calculations.

‡ Formic acid levels are assumed to be initially zero.

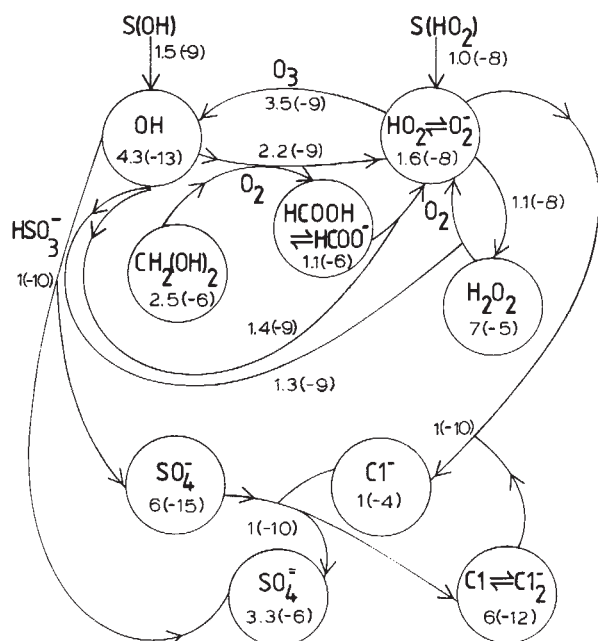


Fig. 1 The primary chemical pathways for aqueous-phase, free radicals in a remote marine cloud. SOH and SHO₂ represent sources of OH and HO₂ to cloud droplets as a result of scavenging from the gas phase. The circles represent reservoirs of dissolved species and the arrows reactive pathways with reactants indicated along the arrow. Model calculated concentrations and reaction rates, in units of M and M s⁻¹, respectively for a remote cloud at T = 295 K and time = 15 min are indicated. Note that 1(-10) = 1 × 10⁻¹⁰.

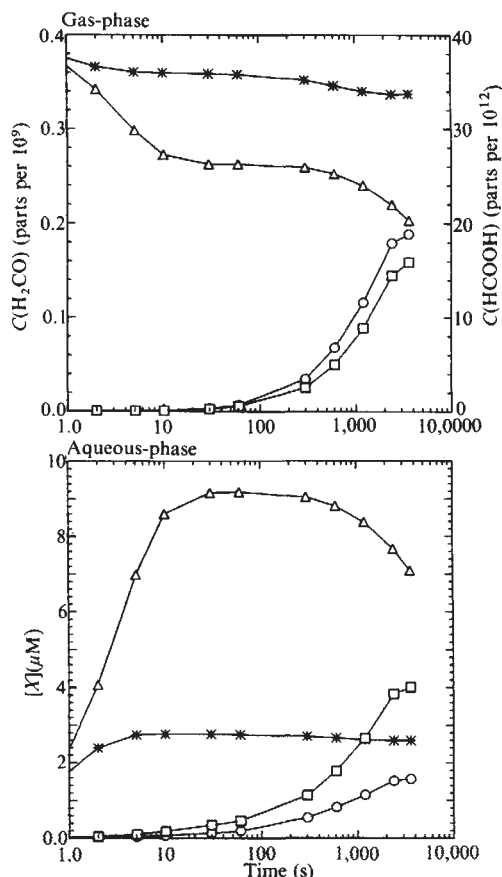
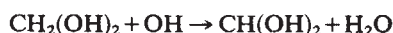


Fig. 2 Gas- and aqueous-phase formaldehyde at temperatures 295 K (*) and 275 K (Δ). Gas- and aqueous-phase formic acid at temperatures 295 K (O) and 275 K (□).

However, CH₂(OH)₂ is not a chromophore and thus should not photolytically decompose. The more probable fate of dissolved formaldehyde is oxidation by OH (ref. 20)



$$k_6 = 2 \times 10^9 \text{ M}^{-1} \text{ s}^{-1} \quad (6)$$

to form the CH(OH)₂ radical. Note that reactions of CH₂(OH)₂ with non-free-radical oxidants such as H₂O₂ appear to be too slow to compete with reaction (6)²¹. In cloudwater where [O₂] > 10⁻⁴ M, CH(OH)₂ reacts predominantly with O₂ to form hydrated formic acid, (HCOOH)_{aq}, through²¹

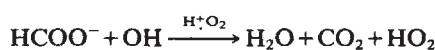


$$k_7 = 4.5 \times 10^9 \text{ M}^{-1} \text{ s}^{-1} \quad (7)$$

For pH > 4 most of the (HCOOH)_{aq} dissociates to form the HCOO⁻ ion¹⁷, for example

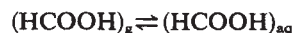


These two species can then be lost from the cloud droplet either by oxidation by OH radicals^{16,22}



$$k_9 = 3 \times 10^9 \text{ M}^{-1} \text{ s}^{-1} \quad (9)$$

or by evaporation, as (HCOOH)_{aq} equilibrates with gaseous formic acid, (HCOOH)_g, such that^{17,18}



$$K_{a,10} = 3.7 \times 10^3 \exp [5,700(1/T - \frac{1}{298})] \text{ M atm}^{-1} \quad (10)$$

The chain oxidation by OH is closed when the HO₂ produced by reaction (7) and reaction (9) is converted back to OH through reactions (1) plus (3) thereby replacing the OH lost in reactions (6) and (9). Chain termination can occur when HO₂ is converted to H₂O₂ (that is reaction (2)) or OH undergoes a reaction such as (4).

To determine the formic acid levels that can be supported by the formaldehyde/formic acid oxidation scheme, we simulated the gas- and aqueous-phase chemistry of a cloud in the remote troposphere with a numerical model. Included in the model's chemical mechanism were the gas-phase reactions of the CH₄-CO-NO_x-O₃-H₂O₂ system^{23,24} and the aqueous-phase reactions previously described¹³, as well as reactions (5) to (10) which were not considered previously. Since the temperature of a cloud is quite variable, calculations were carried out at both 295 and 275 K. For those aqueous-phase rate constants for which only room temperature data are available, reasonable temperature dependencies based on estimated values for their activation energies were assumed.

Adopting the initial conditions in Table 1, the levels of all gas- and aqueous-phase species were calculated by integrating a coupled set of equations of the form of

$$\frac{dn(X)}{dt} = P_g(X) - D_g(X) - N\phi_a(X) \quad (11)$$

and

$$\frac{d[Y]}{dt} = P_{\text{aq}}(Y) - D_{\text{aq}}(Y) + \frac{N\phi_a(Y)}{(4/3\pi a^3) 6.02 \times 10^{23}} \quad (12)$$

where *n* is the number density of a gas-phase species *X*, [*Y*] is the molar concentration of a dissolved species *Y*, *P* and *D* are the appropriate chemical production and destruction rates, *N* is the concentration of cloud drops all having a radius *a*, and *φ_a*(*X*) is the rate at which *X* is dissolved into a single cloud droplet of radius *a*. Following Fuchs and Sutugin²⁵,

$$\Phi_a(X) = \frac{4}{3}\pi lV(X) \frac{a}{1 + \left(0.7 + \frac{4(1-\alpha)}{3\alpha}\right) \frac{l}{a}} \times [n(X) - [X]/K_s(X)RT] \quad (13)$$

where l is the mean free path, V the thermal velocity of X , α the accommodation coefficient assumed to be 0.01 for all soluble species¹³, $K_s(X)$ the solubility constant of X in M atm^{-1} , and $R = 1.36 \times 10^{-22} \text{ atm cm}^3$ per molecule.

The calculated levels of formaldehyde and formic acid are illustrated in Fig. 2. During the first 30 s $(\text{H}_2\text{CO})_g$ decreases and $\text{CH}_2(\text{OH})_2$ increases as formaldehyde equilibrates between the gas and aqueous phases. On longer time scales both $(\text{H}_2\text{CO})_g$ and $\text{CH}_2(\text{OH})_2$ decrease as $\text{CH}_2(\text{OH})_2$ is oxidized and $(\text{H}_2\text{CO})_g$ is transferred to the droplet to maintain equilibrium. Initially, this formaldehyde decrease is balanced by a formic acid increase. After about 50 min, however, the formic acid increase slows as a stationary state is reached in which the rate of formic acid production from reaction (6) is equal to the rate of loss by reaction (9)

$$[\text{CH}_2(\text{OH})_2][\text{OH}]k_6 = [\text{HCOO}^-][\text{OH}]k_9 \quad (14)$$

Thus in mature cloud droplets it is predicted that

$$\frac{[\text{HCOO}^-]}{[\text{CH}_2(\text{OH})_2]} \approx \frac{k_6}{k_9} \sim 1 \quad (15)$$

Since formic acid is both produced and destroyed by OH, the relative levels of $[\text{HCOO}^-]$ and $[\text{CH}_2(\text{OH})_2]$ become independent of $[\text{OH}]$. The $[\text{HCOO}^-]$ level also becomes independent of the initial gas-phase $(\text{HCOOH})_g$ concentration after the stationary state is reached; thus virtually the same final $[\text{HCOO}^-]$ levels were obtained with an initial $(\text{HCOOH})_g$ concentration of 30 p.p.t.v. (parts per 10^{12} by volume) as were obtained with an initial level of zero.

While $[\text{HCOO}^-]$ is not strongly dependent on $[\text{OH}]$ or the initial $(\text{HCOOH})_g$ concentration it is sensitive to temperature, the initial $(\text{H}_2\text{CO})_g$ level and the liquid water content, W_L . As illustrated in Fig. 2, the peak in $[\text{HCOO}^-]$ ranges from $\sim 1.5 \times 10^{-6} \text{ M}$ for 295 K to $\sim 4 \times 10^{-6} \text{ M}$ for 275 K. Since the reactions of the formaldehyde/formic acid oxidation scheme are essentially diffusion limited, their rate constants do not change appreciably with temperature. Thus, the major effect of decreasing temperature is to increase the H_2CO and HCOOH solubilities, resulting in a higher $[\text{HCOO}^-]$. While $[\text{HCOO}^-]$ is linearly dependent on the initial $(\text{H}_2\text{CO})_g$ concentration it is inversely related to W_L ; a doubling in the initial $(\text{H}_2\text{CO})_g$ causes $[\text{HCOO}^-]$ to double also, and a doubling in W_L causes $[\text{HCOO}^-]$ to decrease by a factor of 1.5.

The calculated $[\text{HCOO}^-]$ concentrations are of the same order of magnitude as those observed in rainwater in remote regions (that is 10^{-6} – 10^{-5} M)⁸. Thus the mechanism we have proposed may be an explanation for the formic acid in precipitation. Similarly, a mechanism involving the dissolution and oxidation of CH_3CHO and the production of dissolved CH_3COOH may also be responsible for the CH_3COOH observed in rain. In addition to generating formic acid in rain, the formaldehyde/formic acid oxidation scheme can also act as a source of $(\text{HCOOH})_g$ through a 'gas-to-particle-to gas' conversion process (that is reactions (5) + (6) + (7) + (10)). For instance, if after 1 h the 275 K cloud modelled here were to evaporate completely, the fate of most clouds, the resulting $(\text{HCOOH})_g$ concentration would be about 65 p.p.t.v.; for 295 K the concentration would be about 35 p.p.t.v. As clouds continually form and evaporate in an air mass, this process would cause ambient $(\text{HCOOH})_g$ levels to rise to ~ 35 – 65 p.p.t.v., the exact value depending on the local $(\text{H}_2\text{CO})_g$ concentration and the characteristics of the clouds formed.

The ambient $(\text{HCOOH})_g$ levels of 35–65 p.p.t.v. predicted here are an order of magnitude smaller than those observed in the southwestern US^{10–12}. There are three possible explanations for this apparent discrepancy: (1) the $(\text{H}_2\text{CO})_g$ concentration in the southwestern US is an order of magnitude larger than the value of 0.4 p.p.b.v. (parts per 10^9 by volume) assumed in our model; (2) other sources of formic acid, such as ozone–olefin reactions²⁶, dominate over the aqueous-phase source; or (3) the measurement technique is subject to as yet unidentified experimental errors. Simultaneous measurements of formal-

dehyde and formic acid in both the gas and aqueous phase are needed to help resolve this question.

This work was supported in part by NASA under grant NAG-1-85 and by NSF under grant ATM-8208828.

Received 31 March; accepted 25 May 1983.

1. Eriksson, E. *Tellus* **4**, 215–232 (1952).
2. Barret, E. & Brodin, G. *Tellus* **7**, 251–257 (1955).
3. Cogbill, C. V. & Likens, G. E. *Water. Resour. Res.* **10**, 1133–1137 (1974).
4. Oden, S. *Water, Air, Soil Pollut.* **6**, 137–166 (1976).
5. Penkett, S. A., Jones, B. M. R., Brice, K. A. & Eggleton, A. E. *J. Atmos. Envir.* **13**, 123–137 (1979).
6. Scott, W. D. & Hobbs, P. V. *J. Atmos. Sci.* **24**, 54–57 (1967).
7. Granat, L. *Tellus* **24**, 550–560 (1972).
8. Galloway, J. N., Likens, G. E., Keene, W. C. & Miller, J. M. *J. geophys. Res.* **87**, 8771–8786 (1982).
9. Dawson, G. A., Farmer, J. C. & Moyers, J. L. *Geophys. Res. Lett.* **7**, 725–728 (1982).
10. Chameides, W. L. & Davis, D. D. *Chem. Engng News* **60**, 38–52 (1982).
11. Levy, H. II *Adv. Photochem.* **9**, 364–523 (1974).
12. Logan, J. A., Prather, M. J., Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **86**, 7210–7254 (1981).
13. Chameides, W. L. & Davis, D. D. *J. geophys. Res.* **87**, 4863–4877 (1982).
14. Behar, D., Czapski, G., Rabin, J., Dorfman, L. M. & Schwartz, H. A. *J. phys. Chem.* **74**, 3209–3213 (1970).
15. Bahnmann, D. & Hart, E. J. *J. phys. Chem.* **86**, 252–255 (1982).
16. Farhatziz & Ross, A. B. *Selected Specific Rates of Reactions of Transients from Water in Aqueous Solution* Vol. 3 (NSRDS-DBS 59, NBS Special Publ., 1977).
17. Wagman, D. D. *Selected Values of Chemical Thermodynamic Properties* (NBS Technical Note 270-1, 1965).
18. Weast, R. C. (ed.) *CRC Handbook of Chemistry and Physics* (CRC Press, West Palm Beach, 1979).
19. Galbally, I. E. *J. geophys. Res.* **77**, 7129–7132 (1972).
20. Markovic, V. & Sehested, K. *Proc. 3rd Tihany Symp. on Radiation Chemistry* (eds Dobos, J. & Hedvig, P.) 1269–1280 (Akademai Kiado, Budapest, 1972).
21. Bothe, E. & Schulte-Frohlinde, D. *Z. Naturforsch.* **35b**, 1035–1039 (1980).
22. Nenadovic, M. T., Draganic, Z. D., Draganic, I. G. & Kidric, B. in *Proc. of the 3rd Tihany Symp. on Radiation Chemistry* (eds Dobos, J. & Hedvig, P.) 1269–1280 (Akademai Kiado, Budapest, 1972).
23. Chameides, W. L. & Tan, A. *J. geophys. Res.* **86**, 5209–5223 (1981).
24. Demore, W. B. *et al.* *JPL Publ.* 82–57 (1982).
25. Fuchs, N. A. & Sutugin, A. G. in *International Reviews of Aerosol Physics and Chemistry* Vol. 2 (eds Hidy, G. M. & Brock, J. R.) 1–60 (Pergamon, New York, 1971).
26. Su, F., Calvert, J. G. & Shaw, J. H. *J. phys. Chem.* **84**, 239–246 (1982).

Particle size distributions of n -alkanes and ^{210}Pb in aerosols off the coast of Peru

J. K. Schneider, R. B. Gagosian*, J. K. Cochran & T. W. Trull

Department of Chemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

Lipid class compounds have been used to determine the sources of organic material in atmospheric aerosols from remote marine areas^{1–6}. Recently, it has been shown that the temporal variability of terrestrial source marker organic compounds in aerosols from the tropical North Pacific is directly related to the seasonality of dust storm activity in China^{3,4}. However, little is known about the detailed emission and transport mechanisms of these compound classes from the continents. We report here, as part of the Sea/Air Exchange Program (SEAREX), the particle size distributions of aliphatic hydrocarbons and ^{210}Pb in aerosols in the upwelling zone off the coast of Peru to elucidate these mechanisms.

Cascade impactor samples (Sierra High Volume, Model 235) were collected 10–100 km off the coast of Peru at approximately 15°S during cruise 108, leg 3 of R/V *Atlantis II* during March–April 1981. The main wind direction was between ESE and SSW with an average wind speed of $\sim 10 \text{ m s}^{-1}$. The sampling tower was mounted on the bow of the ship and the aerosol sampler was located on the top of the tower about 15 m above the sea surface. Impactors were operated at $68 \text{ m}^3 \text{ h}^{-1}$ with Sierra Instruments glass fibre filters. The ship was always steaming into the wind during sampling with a sampling sector of $\pm 45^\circ$ of the wind direction. The relative wind speed was $> 5 \text{ m s}^{-1}$.

* To whom correspondence should be addressed.

Three cascade impactor samples were collected and have been analysed for aliphatic hydrocarbons, polycyclic aromatic hydrocarbons, wax esters, fatty alcohols, steroid alcohols and fatty acid salts. One-quarter aliquots of two of these samples were also analysed for ^{210}Pb . We report here only the aliphatic hydrocarbon and ^{210}Pb results. Comparison of the results of the other compound classes with those from the sea-surface microlayer will be published elsewhere⁷. The sampling methods and analytical procedures for the organic compounds have been summarized by Gagosian *et al.*³ and are presented in detail elsewhere⁸. ^{210}Pb was measured by determination of its daughter ^{210}Po ~1 yr after collection. The impactor filters were totally dissolved in HCl , HNO_3 and HF , and ^{210}Po was determined as described in ref. 9. Blank corrections based on filter blanks were <20% for all samples.

The results for the aliphatic hydrocarbons and ^{210}Pb are summarized in Table 1. The aliphatic hydrocarbon fraction in the range $>\text{C}_{20}$ contained almost exclusively *n*-alkanes. Only traces of aliphatic hydrocarbons $<\text{C}_{20}$ were detected due to losses during sampling and analytical procedures for these more volatile compounds. A small 'hump' of an unresolved complex mixture (UCM) of aliphatic hydrocarbons was present in all samples but more prevalent in the smaller particle size fractions.

The total *n*-alkane and ^{210}Pb concentrations for sample PEAS-3 are about double (611 pg m^{-3} and $16.1 \text{ d.p.m. } 10^{-3} \text{ m}^{-3}$, respectively) compared with the concentrations of *n*-alkanes in PEAS-2 and PEAS-4 (277 and 284 pg m^{-3} respectively) and ^{210}Pb in PEAS-2 ($8.1 \text{ d.p.m. } 10^{-3} \text{ m}^{-3}$). This is probably due to the different sampling sectors (Fig. 1) and daily wind conditions. PEAS-3 was taken ~10–30 km offshore while PEAS-2 and PEAS-4 were taken in a 20–100 km sector offshore. The *n*-alkane concentrations of the individual stages of PEAS-3 are also about double the *n*-alkane concentrations of PEAS-2 and PEAS-4, whereas this is only true for the three smallest particle sizes (stages 4, 5 and back-up filter (BUF)) for ^{210}Pb in PEAS-2 (Table 1).

The *n*-alkane concentrations as a function of carbon number for all stages of PEAS-3 are presented in Fig. 2 (distributions for PEAS-2 and PEAS-4 are very similar to PEAS-3). The predominance of *n*-alkanes with an odd number of carbon atoms (for example, C_{27} , C_{29} , C_{31}) indicating land plant wax^{10,11} or soil⁴ origin is clearly seen in all the stages, although it is much more pronounced in the large particle size fractions (stages 1–3). As the particle size decreases (stages 4, 5, and BUF), the even-carbon-numbered *n*-alkane concentrations increase relative to the odd ones. This suggests a contribution from an additional aliphatic hydrocarbon source, probably of anthropogenic origin which typically shows little or no odd-even *n*-alkane carbon predominance¹². This is in agreement with the fact that small but measurable amounts of unresolved complex mixtures (UCM), a characteristic of 'anthropogenic' sources¹², were found on stages 4, 5 and BUF. However, there is the possibility that a small portion of the aliphatic hydrocarbons found are of biogenic 'marine' origin, since there are reports

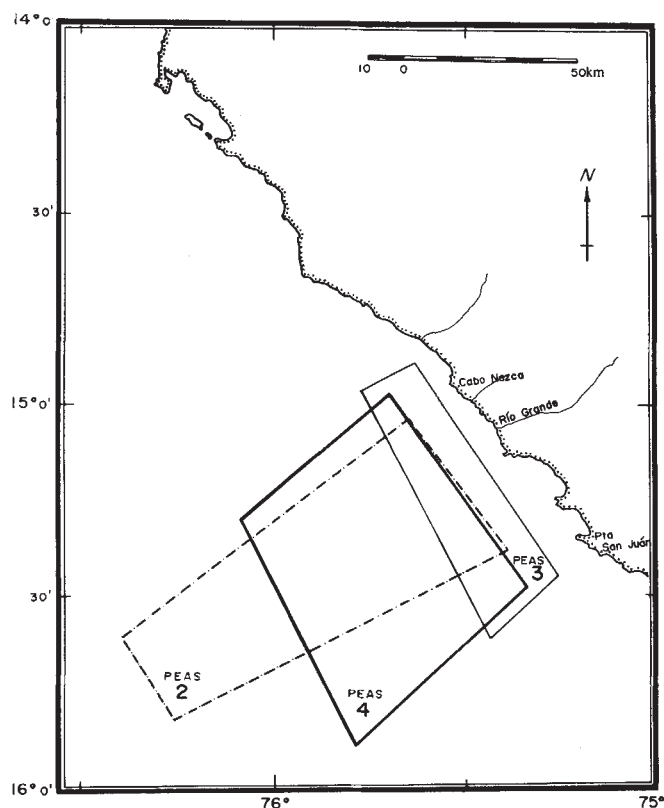


Fig. 1 Sampling site off Peru with sampling sectors for the cascade impactor samples PEAS-2, PEAS-3 and PEAS-4.

of phytoplankton and bacteria producing long-chain *n*-alkanes with no odd-even carbon predominance^{13,14}. Bacteria may also be a source for small amounts of UCM hydrocarbons¹². A third possible source for the *n*-alkanes with a carbon preference index (CPI) = 1 is small particles originating from continental weathering of ancient sediments¹² although this source is considered to be small for the area studied.

The CPIs ($\text{CPI} = \Sigma_{\text{odd}} / \Sigma_{\text{even}}$ for $>\text{C}_{20}$ *n*-alkanes) of the three samples are presented in Fig. 3 as a function of particle size. The shapes of the curves suggest at least two different overlapping *n*-alkane contributions with different particle size distributions. The same general trend in the CPI was observed by Marty and Salot in cascade impactor samples taken in the Gulf of Guinea¹⁵.

Using a very simplified model, the contributions from two sources of *n*-alkanes as a function of particle size can be calculated. The two *n*-alkane sources are a land plant wax or soil source (CPI = 5, approximately the highest CPI value found in these samples and other remote marine samples^{1–4}, representative of terrestrial material emitted into the atmosphere), and

Table 1 Concentrations of the *n*-alkanes and ^{210}Pb as a function of particle size for the cascade impactor samples PEAS-2, PEAS-3, and PEAS-4

	Equivalent aerodynamic diameter* (μm)	<i>n</i> -Alkanes ($\Sigma \text{C}_{21}\text{--C}_{36}$) (pg m^{-3})			$^{210}\text{Pb}^\dagger$ ($\text{d.p.m. } 10^{-3} \text{ m}^{-3}$)	
		PEAS-2	PEAS-3	PEAS-4	PEAS-2	PEAS-3
Mid-date (1981)		17 March	24 March	31 March	17 March	24 March
Volume (m^3)		3,400	3,400	4,000	3,400	3,400
Stage 1	>7.2	24	51	23	0.27 ± 0.03	0.20 ± 0.02
2	>3.0	52	104	47	0.37 ± 0.03	0.39 ± 0.04
3	>1.5	22	69	37	0.40 ± 0.04	0.49 ± 0.03
4	>0.96	51	125	57	1.17 ± 0.07	2.49 ± 0.10
5	>0.50	56	133	70	2.87 ± 0.10	5.53 ± 0.18
BUF	<0.50	72	129	50	3.04 ± 0.14	6.95 ± 0.34
Total		277	611	284	8.1	16.1

* According to the manufacturer, equivalent aerodynamic diameter cutoffs at 50% collection efficiency for particles with a density of 1 g cm^{-3} .

† Uncertainties are based on 1σ counting statistics.

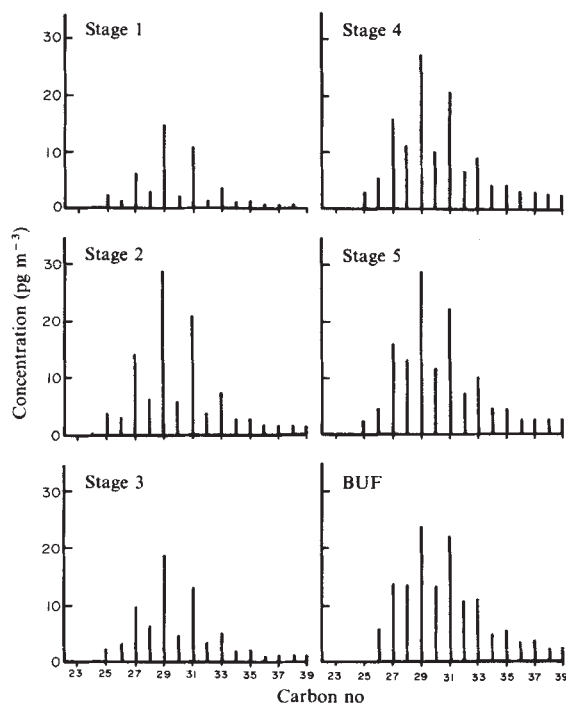


Fig. 2 Concentrations of individual *n*-alkanes as a function of carbon number in the different stages of the cascade impactor sample PEAS-3 (equivalent aerodynamic diameter: stage 1, $>7.2 \mu\text{m}$; stage 2, $>3.0 \mu\text{m}$; stage 3, $>1.5 \mu\text{m}$; stage 4, $>0.96 \mu\text{m}$; stage 5, $>0.50 \mu\text{m}$; BUF, $<0.50 \mu\text{m}$).

an 'anthropogenic' source (CPI = 1). The terrestrial plant wax or soil contribution is relatively evenly distributed over the whole particle size range with a small maximum at stage 2 (Fig. 4a). A very similar distribution pattern was found for other land plant wax compounds such as wax esters and fatty alcohols⁷. This even distribution over different particle sizes is probably due to the wide spectrum of particle sizes generated by epicuticular plant wax and soil aerosol sources, little modified by short-range transport offshore¹⁶. However, the terrestrial biogenic *n*-alkane particle size distribution would be expected to change as a function of distance from land further offshore as observed by Chesselet *et al.*¹⁷ for particulate organic matter. In contrast to the terrestrial *n*-alkanes, the anthropogenic *n*-alkanes show an increase in concentration as the particle size decreases. Applying the same two source *n*-alkane model to Marty and Salot's data¹⁵, we find the same general trends. Increasing concentration with decreasing particle size is observed for ^{210}Pb . This trend is consistent with other studies where it has been shown that ^{210}Pb is mostly present in the smallest size fraction of aerosols over land and in the marine atmosphere^{18,19}. ^{210}Pb derives ultimately from a gaseous land source (^{222}Rn emanating from rocks and soils) but is formed and attaches to small particles in the atmosphere.

There are two possible explanations for the similar trends of ^{210}Pb and the anthropogenic *n*-alkanes as a function of particle size (Fig. 4b,c). First, the distribution pattern of the anthropogenic *n*-alkanes could be due to emission of small particles containing *n*-alkanes during fossil fuel combustion processes^{16,20,21} or other anthropogenic activities. A second explanation involves gas-particle partitioning of the *n*-alkanes. Gas phase *n*-alkanes (for example, evaporated during fossil fuel combustion) can adsorb or condense onto the small carbon particles produced from anthropogenic activities. These alkanes can also be incorporated into or attached to particles covered with natural organic substances. The fact that gas phase *n*-alkanes from different locations exhibit low CPI values similar to those of the anthropogenic *n*-alkanes found in this study²²⁻²⁸, but the CPI for the corresponding aerosol samples vary and

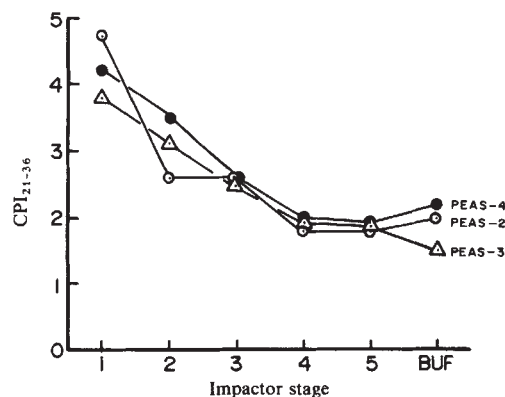


Fig. 3 Carbon preference indices (CPIs) as a function of particle size in the cascade impactor samples PEAS-2, PEAS-3 and PEAS-4 (CPI calculated for the range $\text{C}_{21}\text{--C}_{36}$).

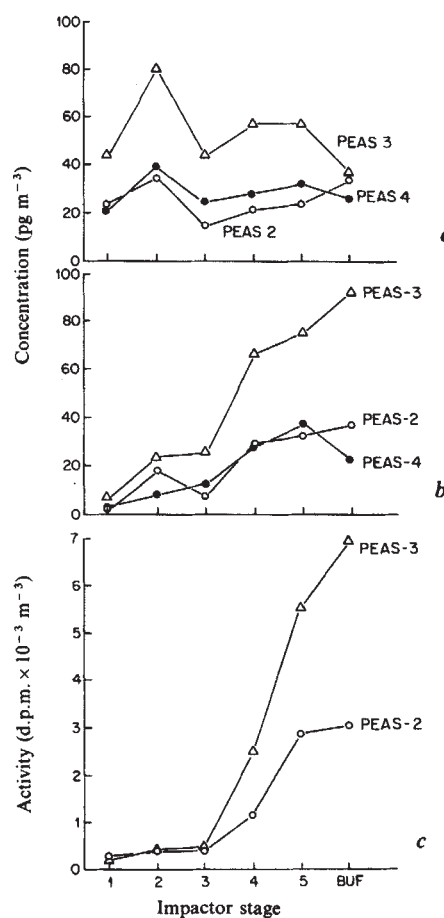


Fig. 4 Calculated concentrations for two proposed *n*-alkane contributions and ^{210}Pb as a function of particle size for the cascade impactor samples PEAS-2, PEAS-3 and PEAS-4: a, terrestrial *n*-alkanes; b, 'anthropogenic' *n*-alkanes; c, ^{210}Pb .

show from little to strong influence of land plant wax *n*-alkanes (high CPI values) is consistent with the gas-particle partitioning hypothesis. However, at present we cannot differentiate between these two explanations.

Although the sources of the anthropogenic *n*-alkanes are quite different from that of ^{210}Pb , a correlation of both ^{210}Pb and anthropogenic *n*-alkanes with small particle sizes would be expected. This suggests that their atmospheric transport into the open ocean on similar size particles proceeds via similar pathways.

Our results show that the aerosol *n*-alkanes consist of contri-

butions from at least two different sources, the land plant wax *n*-alkanes with a relatively even particle size distribution, and the anthropogenic *n*-alkanes which correlate with ^{210}Pb with increasing concentration as a function of decreasing particle size. These results strongly suggest different emission and transport mechanisms for these two *n*-alkane sources.

We thank the Ocean Sciences Division of the NSF for support of this work as part of the SEAREX programme (OCE 77-12914 and OCE 81-11947). The ^{210}Pb work is part of a cooperative Yale-WHOI Searex effort and is supported by NSF grant

OCE 81-13153 (K. K. Turekian, Principal Investigator) through a subcontract to J.K.C. J.K.S. was supported by the Swiss National Science Foundation (82-903-081) and T.T. was supported by a WHOI-MIT Joint Program fellowship. We thank P. B. Comita, J. W. Farrington, E. T. Peltzer, S. G. Wakeham, and O. C. Zafriou for valuable comments and J. B. Alford for laboratory assistance, N. M. Frew for mass spectral data, and the officers and crew of the R/V *Atlantis II* for a successful cruise. Woods Hole Oceanographic Institution Contribution no. 5349.

Received 15 March; accepted 5 May 1983.

1. Simoneit, B. R. T. *Mar. Chem.* **5**, 443-464 (1977).
2. Simoneit, B. R. T., in *Advances in Organic Geochemistry* (eds Douglas, A. G. & Maxwell, J. R.) 343-352 (Pergamon, Oxford, 1980).
3. Gagosian, R. B., Peltzer, E. T. & Zafriou, O. C. *Nature* **291**, 312-314 (1981).
4. Gagosian, R. B., Zafriou, O. C., Peltzer, E. T. & Alford, J. B. *J. geophys. Res.* **87**, 11, 133-144 (1982).
5. Marty, J. C., Saliot, A., Buat-Menard, P., Chesseelet, R. & Hunter, K. A. *J. geophys. Res.* **84**, 5707-5716 (1979).
6. Barbier, M. D., Tusseau, D., Marty, J. C. & Saliot, A. *Oceanol. Acta* **4**, 77-84 (1981).
7. Schneider, J. K. & Gagosian, R. B. Preprint (Woods Hole Oceanographic Institution, 1983).
8. Peltzer, E. T., Alford, J. B. & Gagosian, R. B. *Woods Hole Ocean. Tech. Rep.* (in the press).
9. Turekian, K. K. & Cochran, J. K. *Nature* **292**, 522-524 (1981); **294**, 670 (1981).
10. Eglinton, G. & Hamilton, R. J. in *Chemical Plant Taxonomy* (ed. Swain, T.) 187-217 (Academic, New York, 1963).
11. Tulloch, A. P., in *Chemistry and Biochemistry of Natural Waxes* (ed. Kolattukudy, P. E.) 236-252 (Elsevier, New York, 1976).
12. Farrington, J. W. *Adv. Chem. Ser. No. 185*, 1-22 (1980).
13. Volkman, J. K., Farrington, J. W., Gagosian, R. B. & Wakeham, S. G., in *Advances in Organic Geochemistry 1981* (in the press).
14. Volkman, J. K., Johns, R. B., Gillan, F. T., Perry, G. J. & Bavor, H. J. Jr *Geochim. cosmochim. Acta* **44**, 1133-1143 (1980).
15. Marty, J. C. & Saliot, A. *Nature* **298**, 144-147 (1982).
16. van Vaec, L. & van Cauwenberghe, K. *Atmos. Envir.* **12**, 2229-2239 (1978).
17. Chesseelet, R., Fontugue, M., Buat-Menard, P., Ezat, U. & Lambert, C. E. *Geophys. Res. Lett.* **8**, 345-348 (1981).
18. Buat-Menard, P. in *NATO Advanced Study Institute in Air-Sea Exchange of Gases and Particles* (ed. Slinn, G.) (in the press).
19. Sanak, J., Gaudry, A. & Lambert, G. *Geophys. Res. Lett.* **8**, 1067-1069 (1981).
20. Chan, T. L. & Lawson, D. R. *Atmos. Envir.* **15**, 1273-1279 (1981).
21. Katz, M. & Chan, C. *Envir. Sci. Technol.* **14**, 838-843 (1982).
22. Eichmann, R. et al. *Atmos. Envir.* **13**, 587-599 (1979).
23. Eichmann, R. et al. *Atmos. Envir.* **14**, 695-703 (1980).
24. Cautreels, W. & van Cauwenberghe, K. *Atmos. Envir.* **12**, 1133-1141 (1978).
25. Doskey, P. V. thesis, Univ. Wisconsin (1982).
26. Zafriou, O. C., Gagosian, R. B., Peltzer, E. T. & Alford, J. B. *J. geophys. Res.* (submitted).
27. Peltzer, E. T., Gagosian, R. B., Zafriou, O. C. & Alford, J. B. in *SEAREX Newslett.* **5**, No. 3, 16-25 (1982).
28. Duce, R. A. & Gagosian, R. B. *J. geophys. Res.* **87**, 7192-7200 (1982).

Polymer crystallization by chemical nucleation

R. Legras & J. P. Mercier

Laboratoire des Hauts Polymeres, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

E. Nield

ICI, Research Department, Petrochemicals and Plastics Division, Wilton, Middlesbrough, Cleveland TS6 8JE, UK

Several attempts¹⁻⁵ have been made to control the rate of crystallization and the morphology by the addition of very finely-divided substances which promote abundant nucleation. However, these studies have mostly been carried out on an empirical basis and, except in a few cases (self-seeding⁶, epitaxy²⁻⁵), the mechanism of action of these nucleating agents is poorly understood despite several attempts at modelling^{1,2}. This is particularly so in the case of most technical nucleating agents such as mica, talc and organic salts. We report here that the mechanism of action of organic nucleating agents such as sodium benzoate and its derived salts completely differs from the generally accepted model, at least in the case of polyesters. The nucleating agent reacts as a true chemical reagent with the molten macromolecules and produces ionic end groups which constitute the true nucleating species.

It has recently been discovered⁷ that organic nucleating agents do not behave as heterogeneous substrates in lowering the energy barrier towards crystal nucleation but dissolve in the polymer and react as true chemical reagents with the molten macromolecule to form the nucleating species. The experimental background that led to this mechanism can be illustrated by reference to the sodium 2-chloro-benzoate/polyethylene terephthalate system.

When a sample of polyethylene terephthalate in the form of a molten film was held at 250 °C under a microscope no birefringence was detected at this temperature. Particles of sodium 2-chloro-benzoate were then put onto the molten film and the temperature was kept constant. The birefringence, which was initially observed after addition of the salt vanished in <1 min. In place of the initially crystalline salt a new birefringence

appeared which was fully developed after 15 min. On cooling, recrystallization occurred but, unlike pure polyethylene terephthalate, the polymer containing a small proportion of salt led to a very fine crystalline texture. Thus sodium 2-chloro-benzoate apparently dissolved in molten polyethylene terephthalate and formed a new species which had a modified nucleation behaviour.

When sodium 2-chloro-benzoate was added to molten polyethylene terephthalate at a temperature between 260 and 300 °C in a Brabender Plastograph, a significant decrease in the melt viscosity was observed. The scission of the polyethylene terephthalate chains, occurring on addition of the salt, was confirmed by measurements of molecular weights of the samples taken from the melt and compared with the pure polymer treated in the same experimental conditions. No change in the shape of the molecular weight distribution and in the heterogeneity index was detected by gel permeation chromatography (GPC). This confirms the statistical character of the attack of the polymer by the salt.

Quenched samples of polyethylene terephthalate/salt mixtures taken from the Brabender Plastograph were examined by differential scanning calorimetry (DSC). The polymer was observed to be well nucleated, the rate of crystallization being substantially greater than that of a polyethylene terephthalate of the same molecular weight in the absence of sodium 2-chloro-benzoate.

So, for polyethylene terephthalate nucleated by 1% of salt and which after reaction has an average molecular weight of 9,200, the half-crystallization time at 110 °C was below 40 s, while for pure polyethylene terephthalate of the same molecular weight, the half crystallization time reached 340 s. This result clearly indicates that the observed acceleration of the crystallization kinetics does not simply result from a decrease in molecular weight.

The process that occurs between polyethylene terephthalate and alkali metal salts is the subject of further detailed studies⁸ but full elucidation of the chemistry involved has been achieved and can be illustrated by the results of model compound studies using dimethyl terephthalate (DMTP) to represent the behaviour of the ethylene terephthalate (PETP) repeat unit as shown in Fig. 1.

Above 200 °C, sodium 2-chloro-benzoate rapidly dissolved in the dimethyl terephthalate melt and, as in the polymer, a

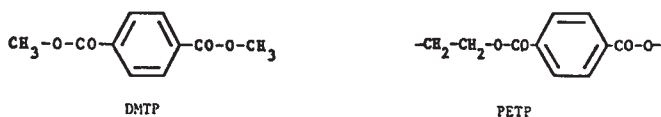


Fig. 1 Structural unit of polyethylene terephthalate (PETP) and of its model compound (DMTP).

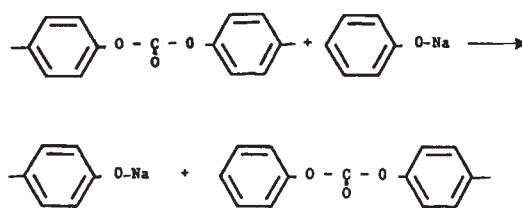


Fig. 3 Attack of the carbonate group by sodium phenate.

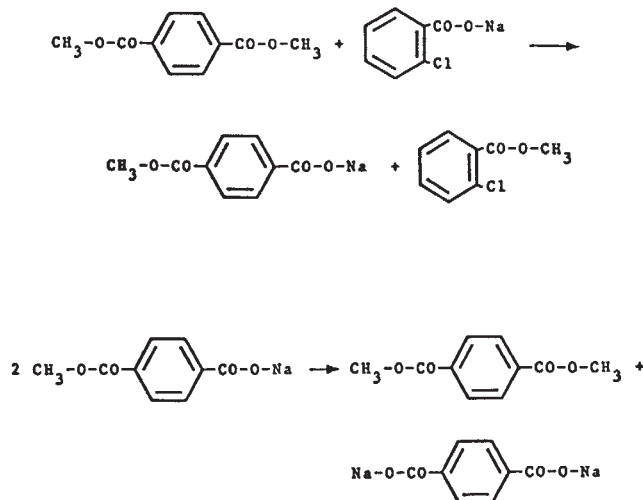


Fig. 2 Reaction path followed by the model compounds above 200 °C.

chemical reaction took place. Analysis by HPLC revealed⁹ the reaction path shown in Fig. 2. Methyl sodium terephthalate formed by reaction of dimethyl terephthalate and sodium 2-chloro-benzoate was insoluble in the reaction medium and the second reaction took place in a separate precipitated phase.

The main products of the reaction of the model compound were also analysed by IR in the 1,500–1,600 cm^{-1} region. The methyl sodium terephthalate had two characteristic absorptions at 1,550 and 1,594 cm^{-1} in the solid state which were shifted to 1,560 and 1,605 cm^{-1} respectively in solution. The disodium terephthalate was characterized by one band at 1,555 cm^{-1} for the solid state and at 1,576 cm^{-1} in solution. The reaction path deduced from model compound studies has been fully confirmed in the polyethylene terephthalate/salt system by IR difference spectroscopy from samples taken from the melt during the mixing process in the Brabender Plastograph⁸.

The spectrum of the sample taken after 1.5 min of reaction at 296 °C disclosed two new bands at 1,550 and 1,594 cm^{-1} respectively similar to that observed in solid methyl sodium terephthalate which suggested that these bands were in fact due to precipitated sodium carboxylate end groups. At a longer reaction time (25 min at 296 °C) further changes were observed that showed that the exchange reaction between sodium carboxylate end groups had occurred with concomitant formation of disodium terephthalate in the molten polymer (appearance of a new absorption at 1,576 cm^{-1} and shift of the band at 1,550–1,555 cm^{-1}). Together with IR modifications, important variations of the molecular weight were observed⁸. The analysis of these data as a function of time and temperature of reaction quantitatively confirmed the reaction mechanism proposed above. Similar behaviour between a range of reactive alkali metal salts and polyethylene terephthalate has been observed.

The bisphenol-A polycarbonate/salt system has not yet been fully characterized¹⁰. Preliminary results indicated that on addition of reactive salts such as sodium 2-chloro-benzoate or sodium phenate, the bisphenol-A polycarbonate melt reacted in a similar way: the attack of the carbonate group with chain scission and formation of ionic end groups is shown in Fig. 3.

The reacted polycarbonate exhibited a very large increase in crystallization rate. For example, pure polycarbonate is charac-

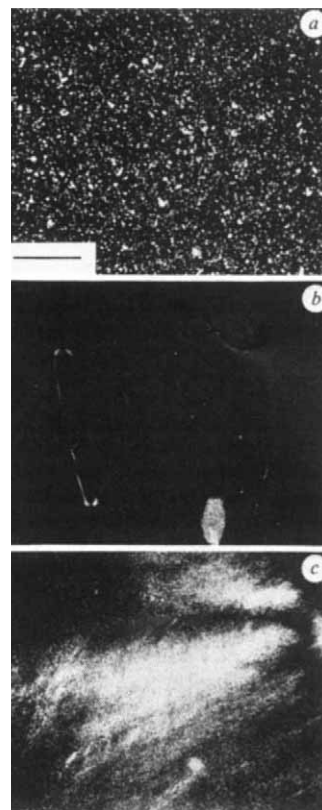


Fig. 4 *a*, Solvent cast polycarbonate film containing 1% of sodium 2-chloro-benzoate. *b*, As *a* after 1 min at 280 °C. The coarse birefringency resulting from the presence of the salt has vanished. *c*, As *b* after 5 min at 225 °C. The very fine birefringency is caused by the polycarbonate crystallization. Scale bar, 200 μm .

terized by a half-crystallization time of 12 days at 190 °C (ref. 11). This value is decreased to below 15 min for a sample nucleated with 0.25% of sodium chlorophenate. This very rapid crystallization occurring after dissolution and reaction of the salt with the polymer melt is illustrated in Fig. 4 for a sample seeded with 1% of sodium 2-chloro-benzoate.

In conclusion, polyesters such as polyethylene terephthalate and bisphenol-A polycarbonate are readily nucleated by alkali metal salts of organic acids, such as carboxylic acids. Contrary to previous claims¹², we have shown conclusively that these salts nucleate through a chemical reaction rather than through a physical effect.

These polymers undergo chain scission on reaction with alkali metal salts to form polymeric species with ionic end groups. This chemical attack was confirmed by IR spectroscopy which has shown in particular that the ionic chain ends are associated into the molten polymer. For that the ionic chain ends are associated into the molten polymer at least in PETP and we believe that the precipitated chain ends form the true nucleating species of the crystallization.

Detailed studies on these and other polymers are continuing with a view to establishing the principles of the 'chemical nucleation' and broadening its applications.

Received 26 February; accepted 16 June 1983.

1. Binsbergen, F. L. *Progress in Solid State Chemistry* Vol. 8 (eds McCaldin, J. O. & Somorjai, G.) 189–238 (Pergamon, Oxford, 1973).
2. Koutsky, J. A., Walton, A. G. & Baer, E. *J. polym. Sci. Polym. Lett.* 5, 177–183 (1967).
3. Mauritz, K. A., Baer, E. & Hopfinger, A. J. *J. polym. Sci. Macromolec. Rev.* 13, 1–61 (1978).
4. Rickert, S. E., Baer, E., Wittmann, J. C. & Kovacs, A. J. *J. polym. Sci. Polym. Phys.* 16, 895–906 (1978).
5. Wittmann, J. C. & Lotz, B. *J. polym. Sci. Polym. Phys.* 19, 1837–1851 (1981); 19, 1853–1864 (1981).
6. Vidotto, G., Lévy, D. & Kovacs, A. J. *Kolloid, Z.* 230, 289–305 (1969).
7. European Patent Application No. 80301860.5 (1980).
8. Legras, R. *et al. Polymer* (in the press).
9. Vanzieghem, A., Dekoninck, J. M., Legras, R. & Mercier, J. P. *J. polym. Sci. Polym. Phys.* (in the press).
10. Phyllis, C., Daumerie, M., Legras, R. & Mercier, J. P. *J. polym. Sci. Polym. Phys.* (in the press).
11. Gallez, F., Legras, R. & Mercier, J. P. *J. polym. Sci. Polym. Phys.* 14, 1367–1377 (1976).
12. Van Antwerpen, F. thesis, Technische Hogeschool Delft (1971).

Constraint on deep mantle viscosity from Lageos acceleration data

W. R. Peltier

Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7

Recently reported analyses of $5\frac{1}{2}$ yr of orbital data for the laser geodynamics satellite (Lageos) have revealed a residual acceleration in the node of its orbit in the amount -14.7 ± 1.3 m arcs yr^{-2} . This acceleration is predominantly attributable to a secular decrease of the degree two component of the Earth's gravitational potential field J_2 . The implied rate of change of J_2 is $(-3.5 \pm 0.3) \times 10^{-11} \text{ yr}^{-1}$. The following analysis argues that this effect is due to Pleistocene deglaciation and that it can be used to constrain strongly the viscosity of the Earth's deep mantle.

The question as to the value of the Earth's viscosity beneath a depth of 670 km remains an important issue. This depth is marked by the presence of a solid–solid phase transition from the low-pressure phase spinel to the high-pressure mixture of the minerals perovskite and magnesiowustite¹. Because this phase transition probably has negative Clapeyron slope and because such transitions appear not to provide any significant impediment to thermal convection through them², it is expected that the mantle convection currents responsible for continental drift and seafloor spreading will be whole mantle in scale.

One physical effect which could undermine this argument in favour of whole mantle convection is the influence of viscosity stratification. If the viscosity of the lower mantle were substantially higher than that of the upper mantle, the circulation could be effectively confined to the upper region. That this situation may in fact exist is a view which has been quite firmly held by some authors.

The purpose here is to provide a geophysical interpretation of a new observational datum which, when combined with previously reported results, seems to rule out the possibility that the viscosity of the lower mantle is significantly higher than that of the upper mantle. This new datum concerns a particular property of the orbit of the artificial Earth satellite Lageos which was launched in 1976 into a nearly circular and drag-free orbit about 1 Earth radius above the surface. Very recently two different analyses have appeared^{3,4} which establish the presence in the time series for the location of the node of the orbit (Ω), a residual secular acceleration after the data have been corrected for known effects.

Of the recently published analyses of $\ddot{\Omega}$, that provided by Yoder *et al.*⁴ is probably the more accurate because it was based on long arcs of tracking data. Two different estimates of $\ddot{\Omega}$ have been obtained based on time scales UT1 derived from both Bureau Internationale de l'Heure (BIH) data and from Lunar Laser Ranging (LLR) observations. The reported accelerations are⁴:

$$\ddot{\Omega}(\text{LLR}) = -14.7 \pm 1.3 \text{ m arcs s yr}^{-2} \quad (1a)$$

$$\ddot{\Omega}(\text{BIH}) = -18.2 \pm 1.6 \text{ m arcs s yr}^{-2} \quad (1b)$$

As it is generally acknowledged that the UT1 based on LLR data is superior to that derived from BIH data, we shall accept equation (1a) as the best current estimate of the residual acceleration. On the basis of the evolution equation satisfied by the keplerian element Ω ⁵, it may be shown that a non-zero value of $\ddot{\Omega}$ implies a time variation of the zonal coefficients J_2 , J_4 , J_6 , and so on in the spherical harmonic expansion of the Earth's gravitational potential field. The dominant contribution is from $\dot{J}_2$ which has been inferred to have the value⁴:

$$\dot{J}_2(\text{LLR}) = (-3.5 \pm 0.3) \times 10^{-11} \text{ yr}^{-1} \quad (2a)$$

$$\dot{J}_2(\text{BIH}) = (-4.4 \pm 0.4) \times 10^{-11} \text{ yr}^{-1} \quad (2b)$$

As a change in J_2 implies a change of the polar moment of inertia C such that $\delta J_2 = (3/2) \delta C / MR^2$, with M and R the Earth's mass and radius respectively, and since the angular momentum $C\omega$ must be conserved in the absence of external torques, the change in J_2 is associated with a change in the Earth's rotation rate ω in the amount $\delta\omega/\omega = -\delta J_2 2MR^2/3C$. Although the dominant contribution to the long time scale change of the Earth's rotation rate is the deceleration due to lunar tidal dissipation in the oceans, it has been possible to extract from historical records of ancient eclipses a nontidal (NT) component in the amount⁶:

$$\left(\frac{\dot{\omega}}{\omega}\right)_{\text{NT}} = (6.9 \pm 2.6) \times 10^{-11} \text{ yr}^{-1} \quad (3)$$

This may be compared with the acceleration expected on the basis of the Lageos observation of $\dot{J}_2$ using the above simple relation between $\delta\omega/\omega$ and δJ_2 . The Lageos data then imply

$$\left(\frac{\dot{\omega}}{\omega}\right)_{\text{NT}}(\text{LLR}) = (7.1 \pm 0.6) \times 10^{-11} \text{ yr}^{-1} \quad (4a)$$

$$\left(\frac{\dot{\omega}}{\omega}\right)_{\text{NT}}(\text{BIH}) = (8.8 \pm 0.8) \times 10^{-11} \text{ yr}^{-1} \quad (4b)$$

It is therefore clear that the newly observed $\dot{J}_2$, based on laser ranging to Lageos, is fully compatible with the nontidal acceleration obtained previously from ancient eclipse observations.

In 1966, Dicke⁷ suggested that the nontidal component of the acceleration of rotation, which seemed to be required to fit the historical eclipse record, was an effect due to Pleistocene

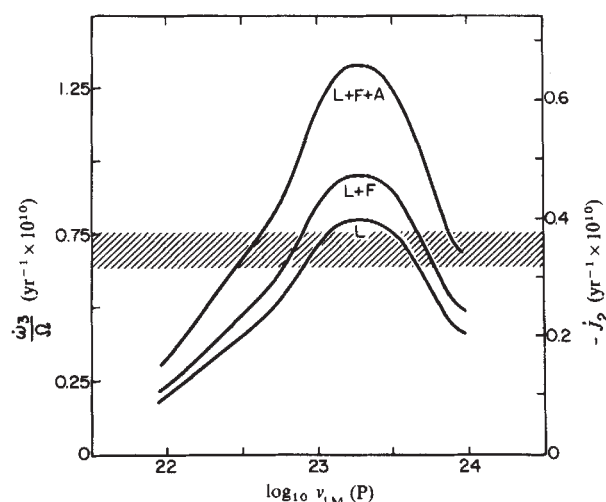


Fig. 1 The variation of the present-day predicted nontidal acceleration $(\dot{\omega}/\omega)_{\text{NT}}$ and the present-day predicted $\dot{J}_2$ as a function of the viscosity of the lower mantle (ν_{LM}) for models with fixed lithospheric thickness, $L = 295.3$ km. Predictions for $L = 120$ km do not differ significantly from these. The hatched region denotes the observed values of these parameters. Calculations are shown for loading by the Laurentian ice sheet only (L), for simultaneous Laurentian and Fennoscandian loading (L+F), and including the Antarctic ice sheet also (L+F+A).

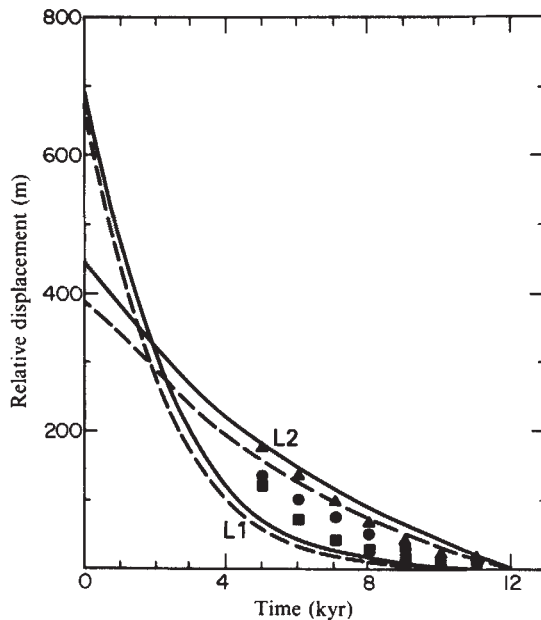


Fig. 2 Time dependence of the relative displacement at the centre of the Laurentian ice sheet for viscosity models L1 and L2. Solid curves are predictions based on the assumption of initial isostatic equilibrium whereas the broken curves include the effect of initial disequilibrium. The relative sea level data points obtained by radiocarbon dating of ancient beaches at Ottawa Islands (■), Churchill (●) and Castle Island (▲) respectively. These sites are all near the centre of the former Laurentide ice sheet.

deglaciation. The last deglaciation event of the current ice age began ~18,000 yr ago. The complete disintegration of the Laurentian (Canadian) and Fennoscandian ice sheets and the partial collapse of those on Greenland and Antarctica led to a global rise in sea level of ~100 m and to a redistribution of matter in the mantle by viscous flow forced by the associated gravitational imbalance. Evidence of the isostatic adjustment (viscous flow) process is clearly revealed in the relative sea level record in the regions which were once ice-covered⁸ (such as the Hudson Bay and Gulf of Bothnia). Dicke and later authors⁹⁻¹¹ used very simple models to invert the observed nontidal acceleration to obtain an average viscosity for the entire mantle. Some of these simple calculations^{9,11} suggested that it might be possible to fit the observation with either of two widely separated values of the viscosity. Here we shall summarize the results of a more sophisticated analysis of the new $\dot{J}_2$ and the old $(\dot{\omega}/\omega)_{NT}$ observations which reveals this ambiguity clearly and shows how it may be resolved.

Table 1 Ice sheet parameters

	Laurentia	Fennoscandia	Antarctica
Mass M_i (kg)	2.0×10^{19}	0.56×10^{19}	0.7×10^{19}
Radius α_i (deg)	15	9.5	20
Colat. α_i (deg)	30	25.5	180
Long. θ_i (deg E)	270	25.0	—

The analysis is based on a new theory of isostatic adjustment⁸ which shows that the predicted values of $(\dot{\omega}/\omega)_{NT}$ and $\dot{J}_2$ are the same to within a constant factor as

$$\dot{J}_2(t) = -\frac{4R^2 \sigma_{20}}{5M I_{33}^R} \left(\frac{\dot{\omega}}{\omega} \right)_{NT} \quad (5)$$

where σ_{20} is the coefficient of second degree and zero order in the spherical harmonic expansion of the load geometry. I_{33}^R is the axial perturbation of inertia which would be associated with the surface load if the Earth were rigid. It is computed from the expression

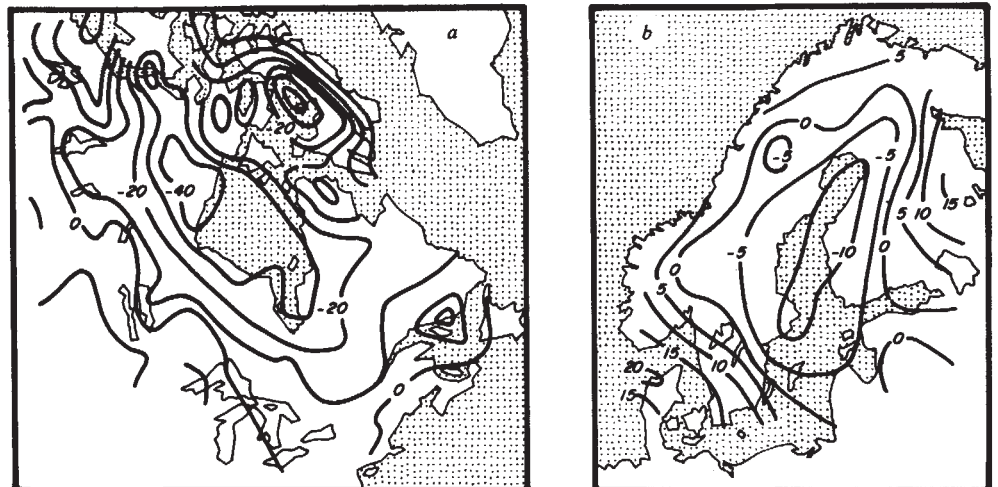
$$I_{33}^R = \sum_{i=1}^N M_i R^2 \left[\frac{2}{15} \frac{a_{20}}{a_{00}} - \frac{1}{3} \cos \alpha_i (1 + \cos \alpha_i) P_2^0(\cos \theta_i) \right] \quad (6)$$

which is obtained assuming that the ice caps are circular and discharge their melt water into ocean basins of realistic shape. M_i , α_i and θ_i are respectively the mass, radius in degrees, and colatitude of the centroid of the i th ice cap, values of which for the main ice masses are listed in Table 1. $a_{20}/a_{00} = -0.1925$ is a ratio of spherical harmonic coefficients of the ocean function used to describe the shape of the ocean basins⁶. The detailed mathematical forms of the predictions for $\dot{J}_2$ and $(\dot{\omega}/\omega)_{NT}$ for cyclic glacial loading are available elsewhere^{8,12}. Here we will restrict discussion to the results.

Figure 1 shows the theoretical predictions of present day $(\dot{\omega}/\omega)_{NT}$ and $\dot{J}_2$ as a function of the viscosity of the Earth's mantle beneath a depth of 670 km with the upper mantle viscosity held fixed at 10^{22} P. Predictions are shown for three different models denoted by L, L+F, and L+F+A which respectively contain one, two and all three ice sheets. Inspection shows that the lower mantle viscosity required by the data is either near 3×10^{22} P or near 10^{24} P. The rotation data cannot themselves be used to distinguish which of these equally plausible values is correct.

There are fortunately two different kinds of geophysical data which can be invoked to remove this ambiguity of interpretation which is intrinsic to the new $\dot{J}_2$ observation. The first of these consists of radiocarbon age-controlled relative sea level data from sites which were under or near the major ice sheets. Figure 2 compares predictions of relative sea level at the centre of a circular disk model of the Laurentian ice sheet with actual relative sea level data from three sites in and around Hudson Bay. Theoretical predictions are shown for models L1 and L2

Fig. 3 Free air gravity anomalies (in mGal) over: *a*, Hudson Bay; and *b*, the Gulf of Bothnia.



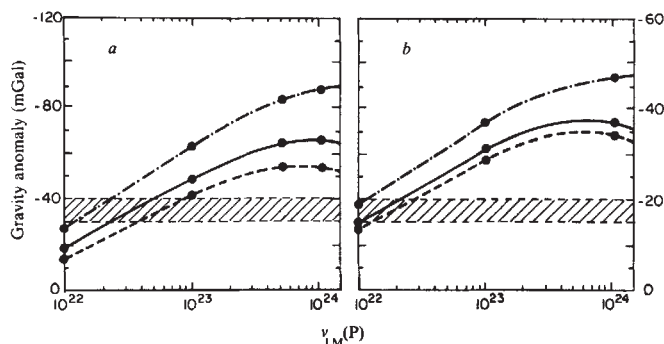


Fig. 4 Predicted present-day peak free air gravity anomaly at the centre of disk model Laurentian (a) and Fennoscandian (b) loads as a function of the viscosity of the mantle beneath 670 km depth. The viscosity of the upper mantle is held fixed at 10^{22} P. The hatched region on each plate denotes the observed peak anomaly in the corresponding region.

which have lower mantle viscosities of 10^{22} P and 10^{23} P respectively and upper mantle values of 10^{22} P. The solid curves are based on the assumption of isostatic equilibrium while the dashed curves include the influence of a load history constrained by $^{18}\text{O}/^{16}\text{O}$ data from deep sea sedimentary cores⁸. Since all the data are bracketed by the predictions of L1 and L2 the implication is that lower mantle viscosity ν_{LM} is such that $10^{22} \text{ P} \leq \nu_{LM} \leq 10^{23} \text{ P}$. These sea level data therefore exclude the higher of the two values of ν_{LM} allowed by the $\dot{J}_2$ observation and from Fig. 1 we have

$$2.7 \times 10^{22} \text{ P} \leq \nu_{LM} (\text{Lageos}) \leq 4.4 \times 10^{22} \text{ P} \quad (7)$$

One further piece of observational evidence which may be used to confirm equation (7) consists of surface free air gravity anomalies over the main centres of Pleistocene deglaciation. Figure 3 shows free air gravity maps for Hudson Bay and the Gulf of Bothnia. The peak free air anomaly over the former region is near -40 mGal whereas that over the latter, appropriately corrected for large-scale bias, is near -17 mGal (ref. 8). These observed peak anomalies are shown as the cross-hatched regions on Fig. 4 where they are compared with theoretical predictions of models which differ from one another only in their values of ν_{LM} . Focusing attention on the solid lines on each plate, which are the most accurate predictions^{12,13}, we note that the value of the lower mantle viscosity required by the free air gravity data is in accord with that inferred from the new $\dot{J}_2$ observation.

The fact that the ν_{LM} inferred from Lageos tracking data agrees with that required by local gravity anomalies suggests that the viscosity of the lower mantle may not have appreciable large-scale heterogeneity. This is expected if this region of the Earth is well mixed by convection. Also, the observed increase of the viscosity across the seismic discontinuity at 670 km depth is most easily understood if the temperature change at this depth is adiabatic¹⁴. This situation will hold only if the circulation is whole mantle in scale and mass transport through the 670-km phase boundary is unimpeded.

Unfortunately, this argument for whole mantle convection is not definitive. We are not yet able to exclude the possibility of a layered convective circulation with a sharp thermal boundary layer at 670 km depth. In this scenario, the rapid increase of temperature through the boundary layer would have to be accompanied by a rather large increase of creep activation energy¹⁴ in order that the viscosity of the lower mantle remain near the viscosity of the upper mantle as required by the new $\dot{J}_2$ observation. This model has the additional interesting feature that there would exist a thin layer of very low viscosity at the base of the upper mantle on the spinel side of the 670-km phase transformation. Work is in progress to determine whether the presence of such a layer can be ruled out on the basis of the relative sea level, free air gravity, and rotation data of

postglacial rebound. The new $\dot{J}_2$ observation from Lageos laser ranging data is an important addition to the data set which we are using in a study of the inverse problem for the viscosity of the Earth's mantle.

Received 25 February; accepted 13 June 1983.

1. Yagi, T., Bell, P. M. & Mao, H. K. *Yb Carnegie Instn Wash.* **78**, 614-618 (1979).
2. Richter, F. M. *Geophys. J. R. astr. Soc.* **35**, 365-376 (1973).
3. Rubincam, D. P. in *Abstr. 5th A. NASA Geodynamics Program Conf. and Crystal Dynamics Project Review*, 31 (NASA Headquarters, Washington DC, 1983).
4. Yoder, C. F. et al. *Nature* **303**, 757-762 (1983).
5. Kaula, W. M. *Theory of Satellite Geodesy* (Blaisdell, Massachusetts, 1966).
6. Lambeck, K. *The Earth's Variable Rotation: Geophysical Causes and Consequences* (Cambridge University Press, London, 1980).
7. Dicke, R. H., in *The Earth-Moon System* (eds Marsden, B. G. & Camero, A. G. W.) (Plenum, New York, 1966).
8. Peltier, W. R. *Adv. Geophys.* **24**, 1-146 (1982).
9. O'Connell, R. J. *Geophys. J. R. astr. Soc.* **23**, 299-327 (1971).
10. Nakiboglu, S. M. & Lambeck, K. *Geophys. J. R. astr. Soc.* **62**, 49-58 (1980).
11. Sabadini, R. & Peltier, W. R. *Geophys. J. R. astr. Soc.* **66**, 533-578 (1981).
12. Wu, P. & Peltier, W. R. *Geophys. J. R. astr. Soc.* **72**, (in the press).
13. Peltier, W. R. & Wu, P. *Geophys. Res. Lett.* **9**, 731-734 (1982).
14. Peltier, W. R. & Jarvis, G. T. *Phys. Earth planet. Inter.* **29**, 281-304 (1982).

Visual motion and cortical velocity

A. Johnston & M. J. Wright

Department of Psychology, Brunel University, Kingston Lane, Uxbridge, Middlesex UB8 3PH, UK

Recent studies have revealed some remarkably simple relationships between visual performance and the neuroanatomy of the visual pathways. The visual field is mapped topographically on the surface of the striate cortex in man; the projection is large for the central visual field and is progressively compressed towards the periphery. Visual acuity decreases with distance from the fovea in proportion to the estimated cortical magnification factor, M (the extent of striate cortex in millimetres corresponding to a degree of arc in visual space)¹. If a stimulus is magnified at peripheral locations in proportion to $1/M$, it becomes equally resolvable across the visual field²⁻⁵. This scaling procedure (M -scaling) maintains equivalence of the cortical projection of stimuli with different visual field loci. We have used M -scaling to investigate motion perception as a visual field variable. We report here that both the lower threshold of motion and adaptation to motion are uniform for M -scaled stimuli, and are related to the velocity of the 'cortical image'.

The sensitivity to contrast (1/threshold contrast) of the human visual system is usually determined as a function of the spatial frequency (cycles per degree of visual angle) of grating patterns with sine-wave luminance profiles. Contrast sensitivity decreases in a systematic manner as gratings are positioned further into the periphery of the visual field. M -scaling increases the angular subtense and reduces the spatial frequency of grating patches having increasing visual field eccentricity while maintaining constant their cortical spatial frequency (cycles mm^{-1}). Contrast sensitivity functions measured using M -scaled gratings at different positions in the visual field superimpose when plotted in terms of cortical spatial frequency²⁻⁵. These relationships are independent of the rate of flicker or drift of the grating^{4,6} and also apply to colour contrast⁵. M -scaling shows the contrast sensitivity of central and peripheral visual fields to be quantitatively rather than qualitatively different. Qualitative differences may, however, exist when we consider other metrics of visual function, such as vernier acuity⁷.

There have been several recent reports that moving patterns viewed in the periphery can be seen to slow down or stop⁸⁻¹², suggesting major differences in motion thresholds for foveal and peripheral vision which might be explicable in terms of the cortical magnification factor. We decided to investigate this phenomenon, quantifying the rate of movement at which a drifting grating can just be distinguished from a stationary one

Fig. 1 *A*, The lower threshold of motion (LTM) expressed as temporal frequency (ordinate) for gratings of various spatial frequencies (abscissa) measured at four eccentricities: $\diamond$, 0 deg; $\square$, 1.5 deg; $\triangle$, 4 deg; $\circ$, 7.5 deg. *B*, The abscissa has been transformed to provide values in terms of cortical spatial frequency. Regression line fitted by least-squares method. *C*, The ordinate has also been transformed to express the LTM in terms of cortical velocity. Subject A.J.

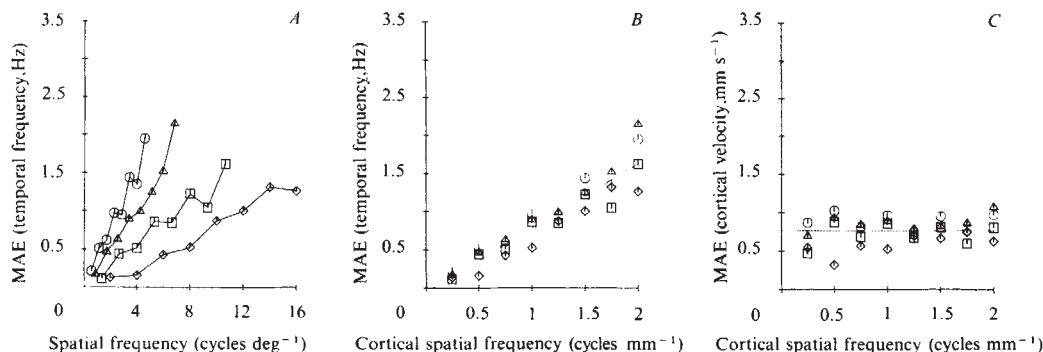
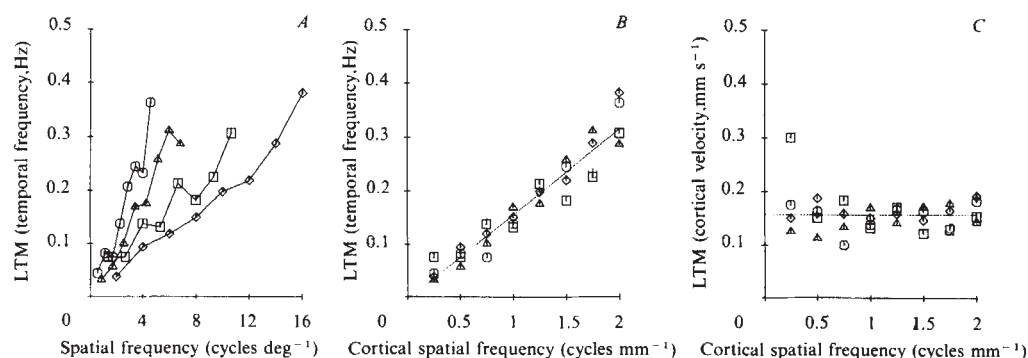


Fig. 2 *A*, Motion aftereffect (MAE) expressed as temporal frequency (ordinate) for gratings of various spatial frequencies (abscissa) measured at four eccentricities: $\diamond$, 0 deg; $\square$, 1.5 deg; $\triangle$, 4 deg; $\circ$, 7.5 deg. Adapt and test spatial frequencies were always the same. *B*, The abscissa has been transformed to provide values in terms of cortical spatial frequency. Regression line fitted by least-squares method. *C*, The ordinate has also been transformed to express the MAE in terms of cortical velocity. Subject M.J.W.

(lower threshold of motion) as a function of spatial frequency at four eccentricities.

A vertical sinusoidal grating was viewed binocularly through a semi-circular aperture in a mask of the same mean luminance (100 cd m^{-2}) at a distance of 200 cm (ref. 6). The aperture had a horizontal base of 7 cm and the grating was fixated at the centre of the base for foveal viewing, or using a series of fixation dots placed vertically below for eccentric viewing. To maintain a constant state of contrast adaptation, a stationary grating (contrast 0.3) was inspected for 30 s at the start of each trial. Following this there were one 6-s and five 4-s periods during which the observer could control the drift rate of this grating by the adjustment of a multi-turn potentiometer. The rate of drift was increased until movement of the grating could just be discerned. Intervals during which adjustments could be made were interspersed with 3-s periods in which a stationary grating was re-instated. A subsequent control experiment using a blank screen of average mean luminance instead of the stationary grating gave similar results. Settings were obtained for 8 spatial frequencies ($SF = 2, 4, 6, 8, 10, 12, 14$ and $16 \text{ cycles deg}^{-1}$) and each data point is an average of thresholds determined for just discernible movement in each of the two horizontal directions of drift. These observations were repeated in identical conditions but with fixation displaced along the lower vertical meridian and viewing distance altered in accordance with M -scaling. Thus, there were four sets of measurements with fixation points 0, 3.5, 6 and 7.5 cm and corresponding viewing distances of 200, 133, 86 and 57 cm so that the lower edge of the aperture fell 0, 1.5, 4 and 7.5 deg from the fixation points. This ensured that each spatial frequency presented decreased with eccentricity such that SF/M (cortical spatial frequency) remained constant. Values of M for these eccentricities have been taken from Virsu *et al.*⁴.

We plotted the magnitude of the lower threshold of motion (LTM) against spatial frequency (Fig. 1*A*) and cortical spatial frequency (Fig. 1*B*) for subject A.J. (The results for M.J.W. were in close correspondence.) LTM, expressed as temporal frequency (number of grating cycles passing a fixed point per second, Hz), increased in an approximately linear fashion with spatial frequency for all eccentricities studied. As velocity = temporal frequency/spatial frequency, we may infer from the data of Fig. 1*A* that LTM at a given visual field locus approximates to a constant velocity for all spatial frequencies. These

data complement the finding¹³ that the threshold for detecting an instantaneous displacement (IDT) for a centrally fixated grating is a constant distance for all spatial frequencies. LTM velocity and IDT are thus both independent of spatial frequency. The LTM measured as a retinal variable increases with eccentric viewing, as expected (Fig. 1*A*). When scaled with respect to the cortical magnification factor, differences in the slopes of these functions disappeared (Fig. 1*B*). Figure 1*C* demonstrates that the LTM is constant in terms of cortical velocity: the average value for A.J. was 0.16 mm s^{-1} and for M.J.W. 0.15 mm s^{-1} . For comparison, Cowey and Rolls¹⁴ found the minimum angle of resolution to be equivalent to a cortical distance of 0.084 mm .

LTM is independent of spatial frequency. Generally, for cells in retina¹⁵ and cortex¹⁶, the optimal spatial frequency is inversely related to receptive field size, therefore the size of receptive fields in retina or cortex does not determine LTM. The increase in LTM considered as a retinal variable correlates with the decrease in cell density with eccentricity found in the retino-cortical mapping¹, but LTM is most economically described as a constant threshold cortical translation velocity.

Barlow¹⁷ has proposed a mechanism whereby spatiotemporal resolution would depend on cell density in the cortex. He argues that cells in the striate cortex can improve on the spatial resolution of retinal ganglion cells by a process of interpolation. Cortical cells would form graded links to groups of input fibres to define the interpolation function. The velocity resolution of striate cortex would also depend on cortical cell density. Assuming uniform cell density across the striate cortex¹⁸ and assuming temporal homogeneity^{4,6}, the velocity resolution of the visual field would vary according to the cortical magnification factor.

Constant cortical velocity thresholds provide a partial explanation of the phenomenon of stopped visual motion. As a grating of a given spatial frequency is moved further into the periphery the cortical image will decrease in size. Eventually, the visual cortex will not be able to resolve small cortical translations within the temporal constraints of the velocity mechanism.

Following prolonged inspection of a pattern moving in a given direction, a stationary pattern will appear to move in the opposite direction. This illusion is the motion aftereffect, MAE (Thompson¹⁹ provides a recent review). In a previous experiment we noted the slowing of gratings when viewed in the

periphery. We thought that greater adaptation to motion in the periphery might add to the stopped motion effect. The MAE was quantified as the temporal frequency of drift required to cancel the illusory perception of motion. Provided fixation was steady, an unambiguous, coherent percept was available at all times and cancellation gave a convincing perception of a steady grating. We adapted for 60 s to a grating (contrast 0.3) drifting at 8 Hz and allowed one 3-s and five 2-s periods in which control over drift rate was passed to the observer, to achieve cancellation of the aftereffect. These adjustment periods were interspersed with 6 s of top-up adaptation. Otherwise, the same experimental conditions were used for MAE as for the LTM determination. Figure 2A–C shows data for M.J.W. (each data point is the average of the MAE in both directions of drift). The MAE measured in terms of cortical velocity was on average 0.78 mm s^{-1} for M.J.W. and 1.2 mm s^{-1} for A.J. This value was constant within the limits of measurement error for different spatial frequencies and eccentricities, but would be influenced by adaptation time and test interval as well as inter-subject

differences. We find the velocities of the MAE to be much greater than the LTM (note the different ordinates for Figs 1 and 2).

There is an interesting paradox in the data. The temporal frequency of the adapt grating was constant, thus its velocity decreased with increasing spatial frequency. Although the velocity of the adapt grating was variable, the velocity of the aftereffect was constant at a given retinal eccentricity and its cortical velocity was constant for all measurements. This may underlie difficulties in determining whether the coding process in motion perception is based on temporal frequency or velocity.

The *M*-scaled properties of the LTM and MAE provide evidence that in the analysis of visual motion, the near periphery is not qualitatively different from foveal vision. The slowing up of peripherally viewed stimuli and the stopped motion effect may be explained by reference to the velocity of the cortical representation.

We thank the MRC for support (grant no. G979/1106/N).

Received 27 April; accepted 22 June 1983.

1. Drasdo, N. *Nature* **226**, 554–556 (1977).
2. Rovamo, J., Virsu, V. & Nasanen, R. *Nature* **271**, 54–56 (1978).
3. Virsu, V. & Rovamo, J. *Expl Brain Res.* **37**, 475–494 (1979).
4. Virsu, V., Rovamo, J., Laurinen, P. & Nasanen, R. *Vision Res.* **22**, 1211–1217 (1982).
5. Noorlander, C., Koenderink, J. J., den Ouden, R. J. & Edens, B. W. *Vision Res.* **23**, 1–11 (1983).
6. Wright, M. J. & Johnston, A. *Vision Res.* (in the press).
7. Westheimer, G. *Vision Res.* **22**, 157–162 (1982).
8. Lichtenstein, M. *J. opt. Soc. Am.* **53**, 302–306 (1963).
9. Campbell, F. W. & Maffei, L. *Nature* **278**, 192 (1979).
10. Campbell, F. W. & Maffei, L. *Vision Res.* **21**, 713–721 (1981).
11. MacKay, D. M. *Perception* **11**, 359–360 (1982).
12. Tynan, P. D. & Sekuler, R. *Vision Res.* **22**, 61–68 (1982).
13. Westheimer, G. *Vision Res.* **18**, 1073–1074 (1978).
14. Cowey, A. & Rolls, E. T. *Expl Brain Res.* **21**, 447–454 (1974).
15. Enroth-Cugell, C. & Robson, J. G. *J. Physiol., Lond.* **187**, 517–552 (1976).
16. Kulikowski, J. J. & Vidyasagar, T. R. *J. Physiol., Lond.* **332**, 10–11P (1982).
17. Barlow, H. B. *Proc. R. Soc. B212*, 1–34 (1981).
18. Hubel, D. H. & Wiesel, T. N. *Proc. R. Soc. B198*, 1–59 (1977).
19. Thompson, P. thesis, Univ. Cambridge (1976).

A functional role for vasoactive intestinal polypeptide in anterior cingulate cortex

James McCulloch & Paul A. T. Kelly

Wellcome Surgical Institute, University of Glasgow, Glasgow G61 1QH, UK

Vasoactive intestinal polypeptide (VIP) is present in high concentrations in the cerebral cortex, where it is the putative neurotransmitter of a major intracortical neuronal system^{1–3}. Homogenates of cortical tissue contain high-affinity, specific binding sites for VIP⁴ as well as an adenylate cyclase system which is sensitive to this peptide^{5,6}. As with many of the other peptidergic systems which have been identified in the central nervous system (CNS), it has proved extremely difficult to elucidate the nature and extent of the functional role of VIP in specific brain areas. Here, using the quantitative autoradiographic ¹⁴C-deoxyglucose technique in rats⁷ to provide insight into functional processes⁸, we describe the increases in glucose utilization which occur locally in anterior cingulate cortex following the unilateral injection of VIP (20 pmol) into this key brain area and, additionally, the focal alterations in glucose use in CNS regions having known neuronal connections with the injected region (for example, ipsilateral mediodorsal thalamus, ventral tegmental area, nucleus accumbens, caudate nucleus and contralateral cingulate cortex). These data provide evidence that VIP may modify the processing of afferent and efferent information within the anterior cingulate cortex in the conscious rat.

The effects of intracingle injections of artificial cerebrospinal fluid (CSF) or VIP (20 pmol) on local cerebral glucose phosphorylation in 31 anatomically discrete regions of the CNS are shown in Table 1. VIP increased glucose utilization in the anterior cingulate cortex and in several CNS regions having known neuronal connections with the injection site⁹, the most notable being in the contralateral anterior cingulate cortex, nucleus accumbens, ventral tegmental areas and mediodorsal thalamic nuclei of both hemispheres (Table 1, Fig. 1). Reduced

glucose utilization was observed following intracortical VIP administration only in that portion of the caudate nucleus (the dorsomedial aspect) (Fig. 2) to which the major fraction of striatal afferents from anterior cingulate cortex are directed⁹. In contrast, glucose use elsewhere in the caudate nucleus was unaltered by the intracortical neuropeptide administration. We emphasize that rates of glucose utilization were similar following cortical injections of VIP or CSF in the overwhelming majority of cerebral areas examined, including several regions having direct neuronal connections with the anterior cingulate cortex (for example, anterior thalamus, amygdala; Table 1). These differential effects of cortical VIP mechanisms in altering activity in only some of the neuronal circuits routed through the anterior cingulate cortex provide further support for a subtle, physiological role for VIP.

Dynamic alterations in glucose utilization seem to reflect predominantly activity in axonal terminals of neuronal pathways¹⁰; and, thus, measurements of focal changes in glucose can be used to identify neuroanatomical circuits which are being activated⁸. The distribution of the alterations of glucose utiliz-

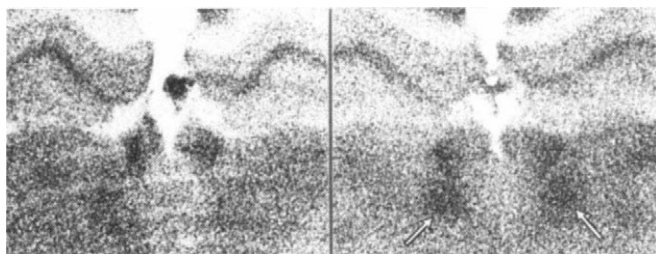


Fig. 1 Representative 2-deoxyglucose autoradiograms at the level of the mediodorsal thalamic nucleus within each animal. Relative rates of glucose utilization are related directly to relative optical density. Following intracingle injection of CSF (left), glucose use in the mediodorsal thalamic nuclei is similar to that of adjacent thalamic areas and lower than that in the lateral habenular nucleus of the epithalamus. Following cingulate injections of VIP (20 pmol) (right), glucose use in the mediodorsal thalamic nuclei (arrowed) is greater than that of adjacent thalamic areas and similar to that in the lateral habenular nuclei.

ation observed after VIP administration into one anterior cingulate cortex, particularly when viewed in relation to the known neuroanatomical connections of the injection site^{9,11}, provides definitive evidence of the neuronal circuits, each with synapses in the cingulate cortex, in which activity can be modified by cortical VIP systems. The anterior cingulate cortex gives rise to a major efferent pathway directed to the contralateral cingulate cortex^{9,11}, and it is presumably changes in activity in this pathway, initiated by VIP, which are reflected as alterations in glucose use in the contralateral cingulate cortex. This major callosal connection between the cingulate cortices provides a neuroanatomical basis for the generally bilateral distribution of the changes in glucose utilization which were associated with unilateral VIP administration. The anterior cingulate cortex projects to and receives projections from the mediodorsal thalamic nucleus^{9,11,12}, ventral tegmental area^{9,12} and nucleus accumbens^{9,13} and, in each of these regions, increases in glucose use resulted from the injection of VIP into the cingulate cortex. In contrast to those regions with reciprocal connections, in the purely orthodromic cingulate projection to the dorsomedial portion of the caudate nucleus, a reduction in glucose use (and cortico-striate activity) was observed.

Many of the regions (anterior cingulate cortex, nucleus accumbens, dorsomedial caudate nucleus) which displayed alterations in glucose utilization also receive dopaminergic projections from the ventral tegmental area¹⁴, which itself displayed

increased glucose use in the present study. A functional interaction between VIP-regulated cingulate efferents and ascending dopaminergic systems is one of several possibilities.

The VIP-mediated increase in phosphorylation of blood-borne glucose within the cingulate cortex may partially underestimate the increase in energy metabolism present as there is *in vitro* evidence that VIP can induce glycogenolysis in mouse cortical slices¹⁵. The breakdown of the small glycogen stores present in the cortex is provoked by VIP in nanomolar concentrations¹⁵ whereas the activation of adenylate cyclase *in vitro* requires micromolar concentrations^{5,6}. The *in vivo* concentration of VIP likely to be achieved by the administration of 20 pmol is in the nanomolar range¹⁶ and would thus stand favourable comparison with previous investigations of VIP mechanisms *in vitro*^{4-6,15}. The ratio of exogenous VIP (20 pmol) injected relative to endogenous VIP in the anterior cingulate cortex (154 pmol per g)² is much lower than similar ratios for intracerebral administration of other polypeptides (for example, neurotensin in the nucleus accumbens^{17,18} and substance P in the ventral mesencephalon^{19,20}) which elicit behavioural alterations. Direct comparisons such as this are fraught with difficulty, as the polypeptides may be distributed differently or inactivated at different rates.

The hypothesis that VIP acts as a neuromodulator in the cerebral cortex is based principally on evidence obtained with conventional immunocytochemical and *in vitro* neurochemical

Table 1 Effects of vasoactive intestinal polypeptide (20 pmol) on local cerebral glucose utilization

	Ipsilateral		Contralateral	
	CSF	VIP	CSF	VIP
Cingulate cortex projections				
Anterior cingulate cortex	72 ± 5	93 ± 4*	78 ± 3	103 ± 5†
Posterior cingulate cortex	70 ± 3	69 ± 2	75 ± 4	78 ± 4
Retrosplenial cortex	64 ± 5	70 ± 5	65 ± 4	75 ± 6
Prefrontal (sulcal) cortex ⁺	101 ± 8	108 ± 10	107 ± 7	113 ± 8
Entorhinal cortex	52 ± 6	51 ± 6	50 ± 5	54 ± 5
Mediodorsal thalamic nucleus ⁺	61 ± 5	85 ± 5†	66 ± 4	90 ± 8*
Anteromedial thalamic nucleus ⁺	83 ± 7	82 ± 3	80 ± 9	92 ± 6
Anteroventral thalamic nucleus	81 ± 9	84 ± 5	85 ± 8	91 ± 5
Lateral habenular nucleus	74 ± 5	84 ± 3	82 ± 6	88 ± 4
Lateral septal nucleus	48 ± 5	53 ± 6	49 ± 5	54 ± 5
Central nucleus of amygdala	37 ± 4	38 ± 2	39 ± 4	39 ± 2
Superior colliculus (deep layers)	57 ± 4	62 ± 5	61 ± 5	63 ± 4
Substantia nigra, compacta (medial) ⁺	47 ± 3	57 ± 4	50 ± 4	57 ± 5
Ventral tegmental area ⁺	40 ± 2	54 ± 4*	39 ± 2	54 ± 5*
Caudate nucleus (dorsomedial)	71 ± 4	57 ± 3*	73 ± 6	66 ± 2
Nucleus accumbens	53 ± 3	70 ± 5*	57 ± 3	73 ± 5*
Limbic system				
Hippocampus (molecular layer)	64 ± 5	65 ± 5	64 ± 5	69 ± 5
Dentate gyrus	50 ± 5	56 ± 4	50 ± 3	56 ± 4
Hypothalamus	43 ± 5	43 ± 2	43 ± 4	43 ± 4
Pontine reticular formation	46 ± 4	43 ± 4	48 ± 4	44 ± 3
Extrapyramidal and motor systems				
Caudate nucleus (lateral)	70 ± 4	73 ± 5	77 ± 3	85 ± 6
Globus pallidus	40 ± 5	42 ± 3	42 ± 4	47 ± 4
Entopeduncular nucleus	41 ± 3	42 ± 5	42 ± 4	46 ± 6
Substantia nigra, compacta (lateral)	49 ± 4	51 ± 5	52 ± 4	56 ± 3
Substantia nigra, reticulata	38 ± 3	41 ± 3	38 ± 3	43 ± 1
Ventrolateral thalamic nucleus	64 ± 7	67 ± 5	70 ± 5	77 ± 6
Motor cortex	77 ± 6	76 ± 6	97 ± 4	108 ± 9
Cerebellar cortex	44 ± 3	41 ± 1	42 ± 3	43 ± 3
Cerebellar nucleus	78 ± 4	79 ± 5	80 ± 5	82 ± 6
Cerebellar white matter	28 ± 1	29 ± 2	30 ± 2	31 ± 2
Vestibular nucleus	88 ± 3	89 ± 5	86 ± 4	90 ± 5

Rates of glucose utilization (μmol per 100 g per min) are presented as mean $\pm$ s.e.m. ($n = 5$ in each group). Glucose utilization was measured in conscious lightly restrained Sprague-Dawley rats (375 g) as described previously^{7,22}. Vasoactive intestinal polypeptide (20 pmol) or a similar volume of artificial CSF (2 μl) (for composition see ref. 23) was administered unilaterally over 2 min via a needle inserted into the anterior cingulate cortex through a stereotactically implanted guide cannula (injection coordinates 0.2 mm L, 2.5 mm V and 1.5 mm anterior to bregma). Both solutions contained Evans blue dye to delineate approximately the distribution of the injectate in each animal. In every animal, the dye was restricted to the anterior cingulate cortex, and no dye was present in the adjacent sensory motor cortex or in subcortical white matter. Measurements of glucose use in the anterior cingulate cortex were not made in the immediate vicinity of the site of injection which could readily be visualized on the autoradiograms. Glucose use in the cingulate cortex was assessed ventral to the injection site. Measurements of glucose utilization were initiated by the intravenous injection of ^{14}C -2-deoxyglucose (125 μCi per kg) 10 min after the intracortical administration of VIP or artificial CSF. + Denotes regions which project to the anterior cingulate cortex as well as receiving efferents from it.

* $P < 0.05$; † $P < 0.01$ for the comparison between VIP and CSF.

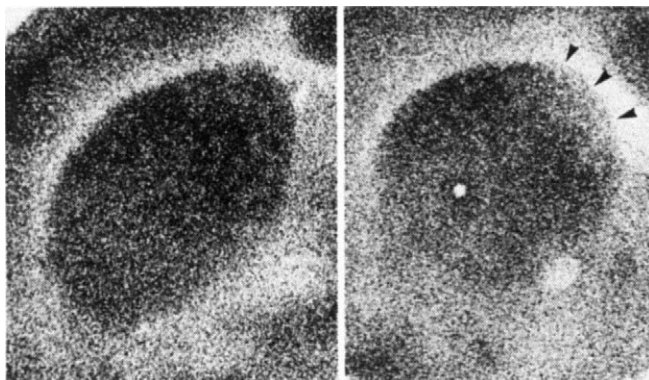


Fig. 2 Representative 2-deoxyglucose autoradiograms of the caudate nucleus, ipsilateral to intracingle injections of CSF (left) or VIP (right). Following CSF injection, glucose use (and optical density) is homogeneous throughout the caudate nucleus, whereas following cingulate injection of VIP, glucose use (and optical density) in the dorsomedial aspect (arrowed) is lower than elsewhere in the caudate nucleus.

approaches¹⁻⁶. The cortex, particularly the cingulate gyrus, contains numerous VIP-immunoreactive neurones whose strictly radial orientation is consistent with the functional organization of the cortex¹⁻³. These neurones are non-spiny bipolar cells whose terminal processes are concentrated in cortical layers I-IV (refs 1-3), and would thus be well placed to modify almost all afferent and efferent activity. However, other than the electrophysiological evidence that some cells in the motor cortex of anaesthetized rats display excitation following iontophoretically applied VIP²¹, there is little evidence from *in vivo* investigations to substantiate the major functional role proposed for cortical VIP systems. We believe that the present studies, particularly the discrete functional alterations provoked by VIP in regions anatomically connected to the cingulate cortex, provide definitive evidence that activity in circuits routed through this cortical region can be modified by VIP mechanisms. In addition, the approach used in the present investigation—stereotactic injections of minute amounts of neuropeptides (pmoles in the case of VIP) and mapping with deoxyglucose autoradiography the local and distant functional consequences in regions with afferent and efferent connections—may prove to be extremely potent in elucidating the extent of the physiological role of various neuropeptides at the wide range of locations where they have been identified immunohistochemically.

These investigations were supported by the Wellcome Trust and the European Science Foundation. The VIP was kindly supplied by Professor V. Mutt. We thank Drs Lars Edvinsson and Rolf Uddman for their encouragement and advice during these studies.

Received 14 March; accepted 23 May 1983.

1. Fuxe, K., Hokfelt, T., Said, S. I. & Mutt, V. *Neurosci. Lett.* **5**, 241-246 (1977).
2. Loren, I. *et al. Neuroscience* **4**, 1953-1956 (1979).
3. Emson, P. C. & Hunt, S. P. in *The Organisation of the Cerebral Cortex* (eds Schmitt, F. O. *et al.*) 325-345 (MIT Press, 1981).
4. Taylor, D. P. & Pert, C. B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 660-664 (1979).
5. Quirk, M., Iversen, L. L. & Bloom, S. R. *Biochem. Pharmac.* **27**, 2209-2213 (1978).
6. Kerwin, R. W., Pay, S., Bhoola, K. D. & Pycoc, C. J. *J. Pharm. Pharmac.* **32**, 561-566 (1980).
7. Sokoloff, L. *et al. J. Neurochem.* **28**, 897-916 (1977).
8. McCulloch, J. *Handbk Psychopharmac.* **15**, 321-410 (1982).
9. Beckstead, R. M. *J. comp. Neurol.* **184**, 43-62 (1979).
10. Schwartz, W. J. *et al. Science* **205**, 723-725 (1979).
11. Vogt, B. A., Rosene, D. L. & Peters, A. *J. comp. Neurol.* **201**, 265-283 (1981).
12. Beckstead, R. M. *J. comp. Neurol.* **166**, 403-416 (1976).
13. Powell, E. W. & Leman, R. B. *Brain Res.* **105**, 389-403 (1976).
14. Fallon, J. H. & Moore, R. Y. *J. comp. Neurol.* **180**, 545-580 (1978).
15. Magistretti, P. J., Morrison, J. H., Shoemaker, W. J., Sapin, V. & Bloom, F. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6535-6539 (1981).
16. McCulloch, J., Kelly, P. A. T., Uddman, R. & Edvinsson, L. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1472-1476 (1983).
17. Ervin, G. N., Birkemo, L. S., Nemeroff, C. B. & Prange, A. J. *Nature* **291**, 73-76 (1981).
18. Emson, P. C., Goedert, M., Horsfield, P., Rioux, F. & St. Pierre, S. *J. Neurochem.* **38**, 992-999 (1982).
19. Eison, A. S. *et al. Science* **215**, 188-190 (1982).
20. Brownstein, M. J., Mroz, E. A., Kizer, S., Palkovits, M. & Leeman, S. E. *Brain Res.* **116**, 299-305 (1976).
21. Phillips, J. W. & Kirkpatrick, J. R. *Can. J. Physiol. Pharmac.* **58**, 612-623 (1980).
22. Kelly, P. A. T. & McCulloch, J. *J. Neurochem.* **39**, 613-624 (1983).
23. McCulloch, J. & Edvinsson, L. *Am. J. Physiol.* **238**, H449-456 (1980).

Independence of the numbers of legs and leg ganglia in *Drosophila bithorax* mutants

Erik Teugels & Alain Ghysen

Laboratoire de Génétique, Université Libre de Bruxelles, 67 rue des Chevaux, 1640 Rhode-St-Genèse, Belgium

A prominent feature of most central nervous systems is the presence of highly organized association centres, often called 'nuclei' in vertebrate brains and 'glomeruli' or 'neuromeres' in invertebrate brains and ventral ganglia. As little is known of the processes leading to the formation of these centres, we have investigated this question in the case of the leg neuromeres of *Drosophila*. The thoracico-abdominal ganglion of wild-type flies contains six conspicuous neuromeres, each associated with a leg¹. In *bithoraxoid* (*bxd*) mutants, the first abdominal segment is transformed to thoracic, and one or two additional legs may develop². We show here that supernumerary neuromeres may also be observed in this mutant. However, in a given individual the number of neuromeres is independent of the number of legs. In *Hyperabdominal* (*Hab*) mutants, the meta-thoracic segment is transformed to abdominal, and the meta-thoracic legs may be missing³. The ganglia of four-legged mutants show a variable reduction of the metathoracic leg neuromeres. We conclude that the formation of a leg neuromere is genetically controlled and does not depend on the presence of a corresponding leg.

In *Drosophila*, all thoracic and abdominal segmental ganglia are fused in a single mass, the thoracico-abdominal ganglion. The only clear trace of segmentation remaining in the adult ganglion is the presence of the three pairs of leg neuromeres. Most of the leg sensory neurones project into the corresponding neuromere⁴, and most of the dendrites of the motoneurones to leg muscles remain confined to this neuromere (ref. 5 and A.G., unpublished observations), thus confirming the idea that each neuromere acts as the association centre for the corresponding leg¹. We have isolated monoclonal antibodies⁶ directed against adult thoracico-abdominal ganglia. When assayed on such ganglia by indirect immunofluorescence, one antibody (1E8) appeared to bind to the three pairs of leg neuromeres, the unpaired abdominal neuromere resulting from the fusion of the eight abdominal neuropiles, and the smaller wing neuromeres (Fig. 1). We have used this antibody to examine the thoracico-abdominal ganglion in mutants of the *bithorax* complex³ that transform abdominal segments to thoracic, or thoracic to abdominal.

The mutation *bxd* transforms the first abdominal segment to thoracic, as evidenced in the adult by the absence of first abdominal epidermis and the occasional appearance of an additional pair of legs as well as, very rarely, supernumerary halteres and metanotum². These mutations have a rather variable expressivity, with phenotypes ranging from no additional leg to two fully developed supernumerary legs. Intermediate phenotypes may show unevolved legs remaining inside the abdomen, or poorly developed legs. We have analysed two alleles, one rearrangement with a breakpoint in the *bxd* region (*bxd*¹⁰⁰) and one point mutant (*bxd*¹). Both were used as hemizygous or heterozygous with *Ubx*⁸⁰, *Ubx*¹⁰⁵, *Df(3)Ubx*¹⁰⁹ or *Df(3)P2*; all combinations gave essentially identical results. The *bxd*¹⁰⁰ mutation is more extreme than *bxd*¹ in that most *bxd*¹⁰⁰ mutant flies do not emerge from the pupal case, while most *bxd*¹ emerge as viable adults.

The thoracico-abdominal ganglion of the extreme mutant, *bxd*¹⁰⁰ (26 flies analysed), showed eight well-defined leg neuromeres (Fig. 2a), even in those individuals where no trace of an additional leg was detected (Fig. 2b). Thus it is clear that an additional leg neuromere can be formed in the absence of a corresponding leg. In the case of the weaker allele, *bxd*¹ (36

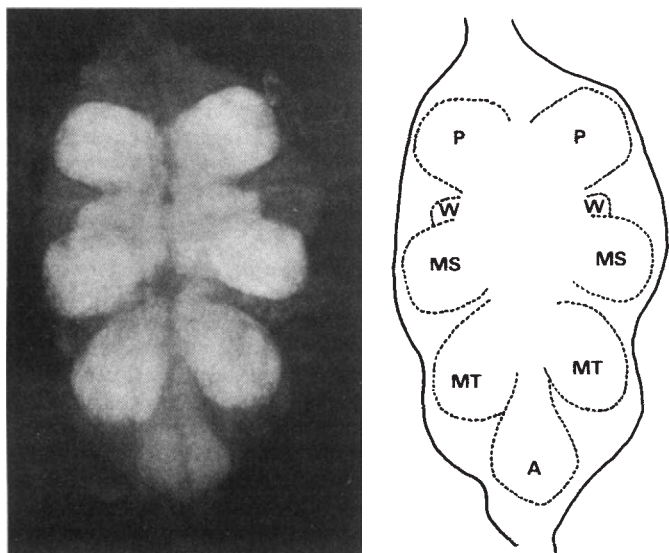


Fig. 1 The neuromeres in the thoraco-abdominal ganglion of *Drosophila*. Left, the thoraco-abdominal ganglion of a wild type fly was dissected in phosphate buffer pH 7.6, fixed for 1 h in acetic acid/ethanol (1:3), rinsed for 30 min in phosphate buffer, incubated for 6 h in supernatants of 1E8 hybridoma cultures diluted 1:4 with phosphate buffer containing 0.5% Triton X-100 and 0.3% deoxycholate, washed for 6 h in the same buffer, incubated for 6 h in fluorescein-coupled goat anti-mouse antibodies (Nordic) diluted 1:40 with the same buffer, washed for 6 h in the same buffer, and finally cleared in glycerol/phosphate buffer (3:1). The photographs were taken on a Zeiss epifluorescence microscope with a $\times 6.3$ neofluar objective. Right, drawing of the outline of the ganglion (solid line) and the various neuromeres (broken line). A, Unpaired abdominal neuromere; MS, MT, P: meso-, meta- and prothoracic leg neuromeres, respectively; W, wing neuromere, also known as the mesothoracic accessory neuromere¹.

flies analysed), no distinct additional neuromere was ever observed (Fig. 2c), even in flies that had eight fully developed legs (Fig. 2d). The reason for this may be that either the mutation *bx^d*¹ is too weak to have a detectable effect on the central nervous system, or this mutation affects preferentially the epidermis. Whatever the reason, this mutant shows that the development of an additional leg is not necessarily accompanied by the formation of a corresponding neuromere. In *bx^d*¹ flies that have eight fully developed legs, the metathoracic leg neuromeres occasionally appeared slightly larger than in wild type (Fig. 2d). This may result from the projection of two complete sets of sensory fibres (from the normal and from the additional legs) causing a swelling of this neuromere. Alternatively, it is possible that this swelling represents an incipient additional leg neuromere which has become closely apposed to the normal metathoracic leg neuromere.

These results show that in *bx^d*¹⁰⁰ flies additional neuromeres are formed even when no additional legs are present, while in *bx^d*¹ the central nervous system shows little if any trace of transformation even when two additional well developed legs are present. From this we conclude that the development of additional neuromeres depends on the genotype of the fly, and not on the presence of additional legs. However, we cannot rule out the possibility that the mutation *bx^d*¹⁰⁰ is affecting some early and possibly crucial input into the developing central nervous system, rather than acting directly on the central neurones.

The mutation *Hab* transforms the last thoracic segment to abdominal³, and therefore reduces the number of legs to four. The penetrance of this mutation is improved in *Hab* flies produced by *Df(3)red^{P93}* mothers (E. Lewis, personal communication), but even then the penetrance is far from complete and the expressivity very variable. Mutants that had six legs always showed six neuromeres, with the metathoracic neuromeres occasionally abnormal in size or shape (25 flies analysed). Among the flies that had only five legs (17 cases) or four legs

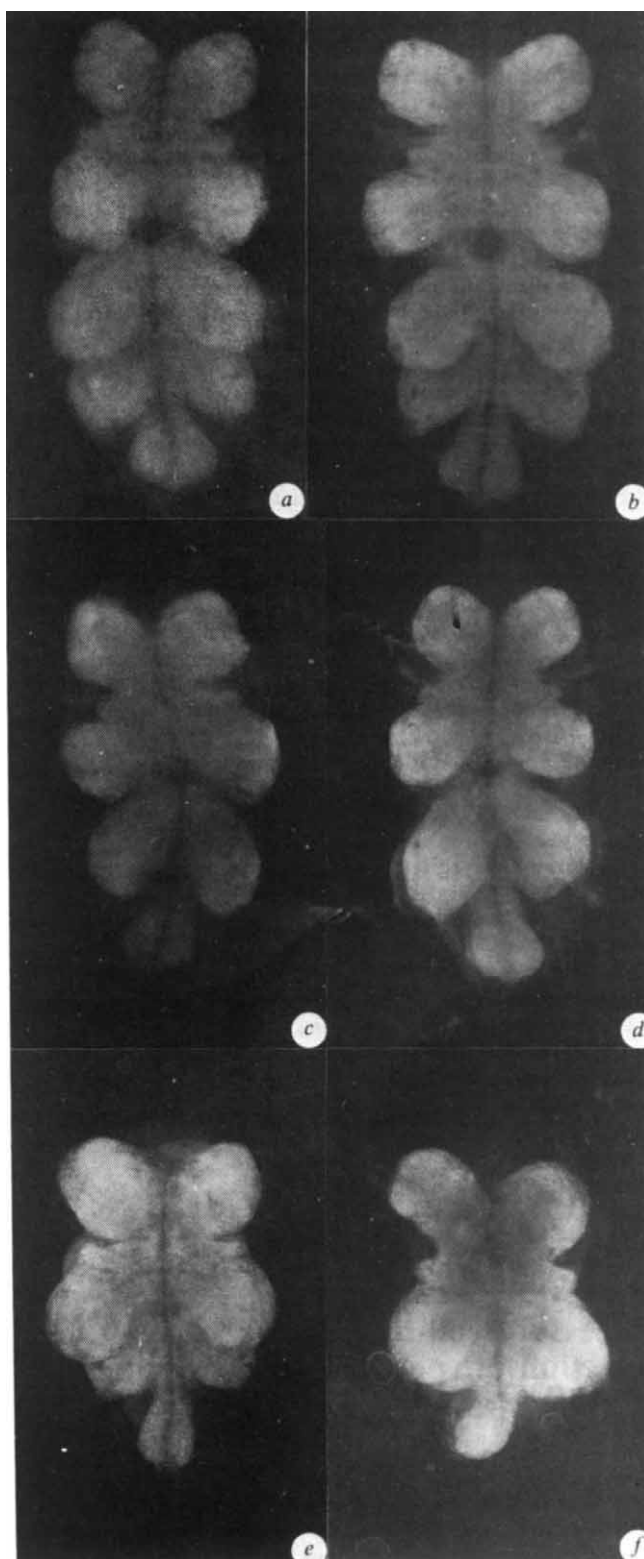


Fig. 2. Effect of various mutations of the bithorax complex on the leg neuromeres. The various alleles used in this work are described in refs 2, 3, 8 and 9. a, Ganglion of a *bx^d*¹⁰⁰/*Ubx*¹⁰⁵ mutant that had one well developed additional leg; b, ganglion of another *bx^d*¹⁰⁰/*Ubx*¹⁰⁵ fly that had no additional legs; c, ganglion of a *bx^d*¹/*Ubx*¹⁰⁵ mutant that had no additional legs; d, ganglion of a *bx^d*¹/*Ubx*¹⁰⁵ fly with two fully developed additional legs; e and f, ganglia of *Hab* mutants that had only four legs. In all *bx^d*¹ and *bx^d*¹⁰⁰ mutant flies that had no recognizable additional legs, we observed a few cuticular vesicles clustered near the junction between the thorax and abdomen. These vesicles were often very small, typically containing 10–50 trichomes; in some flies they were larger and contained one or a few bristles and/or campaniform sensilla.

(18 cases) we found a wide variety of neural phenotypes. In some cases six neuromeres were observed, in other cases one or both metathoracic neuromeres appeared greatly reduced in size (Fig. 2e) and in rare cases they seemed to have almost completely disappeared (Fig. 2f).

These results extend to the normal metathoracic neuromeres the conclusion reached with *bxd* for the additional neuromeres, in that their development depends on the genotype of the fly and not on the presence of normal metathoracic legs.

We draw two conclusions from our results. First, we have shown that mutations in the bithorax complex that transform the metathoracic and first abdominal segments one into another at the level of the epidermis may show a parallel transformation at the level of the central nervous system, producing flies having four or eight neuromeres instead of the normal six (Fig. 2). This indicates that the development of leg neuromeres depends on the process of segmental determination controlled by the bithorax complex, in agreement with the observation that a deficiency for the whole complex has a detectable effect on the embryonic central nervous system^{3,7}.

As the expression of these mutations is not complete, it is possible to analyse the correlation between the epidermal and the neuronal transformation in each individual. We have shown that there is no direct correlation between the presence of a leg and the presence of a corresponding neuromere. More specifically, we have seen cases of an additional neuromere without a corresponding leg (in *bxd*¹⁰⁰), an additional leg without a corresponding neuromere (in *bxd*⁴), and a normal neuromere without a corresponding leg (in *Hab*). Taken together, these observations suggest that the development of a leg neuromere is a normal feature of all segmental ganglia, irrespective of the presence or absence of an associated leg, and is normally repressed in the abdominal ganglia by a function of the bithorax complex.

We thank E. Lewis for unfailing help and R. Thomas for continuous support, and J. D. Franssen for performing the fusion from which the 1E8 hybridoma was obtained and for introducing us to the hybridoma techniques. E.T. is supported by the Fondation van Buuren and A.G. is Chercheur qualifié of the Fonds National de la Recherche Scientifique (Belgium).

Received 13 April; accepted 21 June 1983.

1. Power, M. E. *J. comp. Neurol.* **88**, 347-409 (1948).
2. Lewis, E. B. *Am. Zool.* **3**, 33-56 (1963).
3. Lewis, E. B. *Nature* **276**, 565-570 (1978).
4. Teugels, E. & Ghysen, A. *Prog. Brain Res.* **58** (in the press).
5. Green, S. H. *Nature* **292**, 152-154 (1981).
6. Köhler, G. & Milstein, C. *Nature* **256**, 495 (1975).
7. Jimenez, F. & Campos-Ortega, J. A. *Wilhelm Roux's Arch. dev. Biol.* **190**, 370-373 (1981).
8. Linsley, D. L. & Grell, E. H. *Carnegie Inst. Wash. Publ.* **627** (1968).
9. Capdevila, M. P. & Garcia-Bellido, A. *Wilhelm Roux's Arch. dev. Biol.* **185**, 105-126 (1978).

Accessory cell-dependent selection of specific T-cell functions

G. Ramila & P. Erb

Institute for Microbiology, University of Basel, Petersplatz 10, CH-4003 Basel, Switzerland

Activation of many T-cell functions depends on their interaction with antigen-presenting accessory cells which express I region associated (Ia) products¹⁻⁴. Cells expressing accessory cell function include those of the monocyte/macrophage lineage^{5,6} and dendritic cells^{7,8}. More recently, B cells⁹ and a number of tumour cell lines of macrophage or B-cell origin¹⁰⁻¹⁷ were shown to act as accessory cells in certain assays. We showed previously that normal peritoneal exudate macrophages (PEC) induced both T-cell proliferation as well as T-cell help, whereas various Ia⁺ tumour lines of macrophage or B-cell origin, although stimulating antigen-specific T-cell proliferation, did not significantly activate T-cell help^{18,19}. We report here that during the initial T-cell activation *in vitro* accessory cells (PEC or Ia⁺ tumour cells) select particular T cells to express previously determined functions. Moreover, some tumour cell lines induce suppressor T cells which inhibit helper activity.

The T-cell functions measured in this study were antigen-specific T-cell proliferation, T-cell help for B cells in secondary IgG antibody responses, and T-cell suppression of helper activity. We intentionally used antigen-specific T-cell lines and not T-cell clones for some experiments, first, because they represent a mixture of clones of T cells potentially expressing one or more functions, and second, we wanted to look at events which take place early in T-cell activation.

We first tested whether the functional expression of helper activity was dependent on the proliferative responsiveness of the T cells. Thus, keyhole limpet haemocyanin (KLH)-specific T-cell lines in culture for about 2 weeks were separated according to cell size into three fractions containing cells of large, medium and small size. Aliquots of the separated cells were tested for help and proliferation (Fig. 1a). Among the 2-week-old T-cell lines only the small and medium but not the large T cells expressed significantly helper activity. However, all three fractions proliferated in response to KLH and PEC used as accessory cells. The large-size fraction proliferated best, although a considerable proportion of this proliferation was antigen independent. Remaining cells in the three size fractions were cultured for another 2 weeks and tested again in the same way (Fig. 1b). At the time of the second assay the initially small and medium-size cells had enlarged and thus, there was no longer a size difference. At this time the helper activity pattern was the same as initially observed, that is, the large cells were not effective in expressing helper activity. The small-size cells at the time of the final assay, now large as well, proliferated best, while the proliferative capacity of the two other fractions was lower. Twice more (at 4 weeks and 6 weeks) the three separately cultured fractions were tested and gave similar results. Thus, actively proliferating T cells are not necessarily helper cells. Therefore, the expression of helper activity does not correlate with T-cell proliferative responses. The results of repeated assays of T-cell lines suggested that the initial activation of the T cells by accessory cells was decisive for expression of a particular function, that is, a T cell selected as a helper cell remained a helper cell, while if it was not selected as a helper cell, it did not become one. To test this observation the following experiments were done: KLH-specific T-cell lines were set up using either PEC or BC3A cells (an Ia⁺ leukaemia line) as accessory cells.

After 1 week the cultures were divided and one half of the cells was further incubated with the same accessory cells as before, and the other half with the other accessory cell type (for example, PEC stimulated T-cell lines with BC3A cells and vice versa). Two weeks later the four groups of T-cell lines were tested for help and proliferation. The T cells of all four groups proliferated in response to KLH and either PEC or BC3A cells in a 3-day assay, although the BC3A cells usually generated better proliferation than the PEC (Fig. 2a). However, only the PEC-stimulated T-cell lines provided optimal help to B cells (Fig. 2b). The BC3A cell-stimulated T-cell lines and the T-cell lines stimulated by BC3A followed by PEC did not express helper activity at all. Minimal helper activity was observed in the T-cell lines stimulated first with PEC followed by BC3A cells. Of particular note, the T-cell lines of this group slowly died out, presumably because they were not able to be restimulated by BC3A cells once they were stimulated by PEC. The T-cell lines of the remaining groups were subsequently tested three more times and the same results were seen. Moreover, similar combinations were set up with T-cell lines from other mouse strains and other tumour lines, for example TA3, and identical results were recorded. Thus, using Ia⁺ tumour lines as accessory cells, T-cell lines can be generated which proliferate in response to antigen and Ia⁺ accessory cells but do not express helper activity, in contrast to PEC-stimulated T-cell lines which express both functions.

Similar experiments were performed using *in vivo*-primed T cells instead of T-cell lines, which gave identical results. For T-cell lines as well as for primed T cells we demonstrated that the proliferative response induced by tumour cells is antigen-

specific¹⁸, and moreover, is Ia restricted¹⁹. Antigen-pulsed tumour cells or PEC were also tested to rule out that the antigen dose in conjunction with tumour cells was inappropriate for helper cell activation (Fig. 3). As pulsing was done in excess of antigen and the pulsed BC3A cells induced proliferation, these results show that the tumour cells are not intrinsically

poor presenters. Similar results were obtained if antigen or the tumour cells were titrated over a wide range into the helper cell cultures, that is the tumour lines did not activate helper cells^{18,19}. Walker *et al.*¹⁴ reported that various B lymphoma tumours were highly efficient as accessory cells in T cells proliferation assays but did not support sheep erythrocyte responses. All the tumour lines with accessory cell function described so far have been shown to be effective in allogeneic or antigen-specific T-cell proliferation, in the production of lymphokines by T-cell hybridomas, or in some instances in the reconstitution of the sheep erythrocyte response, but none has yet been reported to activate T helper cells to soluble antigens¹⁰⁻¹⁷.

One possible explanation for the failure of the tumour lines tested to induce T-cell help might be that they activate suppressor mechanisms. We first observed that BC3A cell-stimulated T-cell lines did not provide help after treatment with

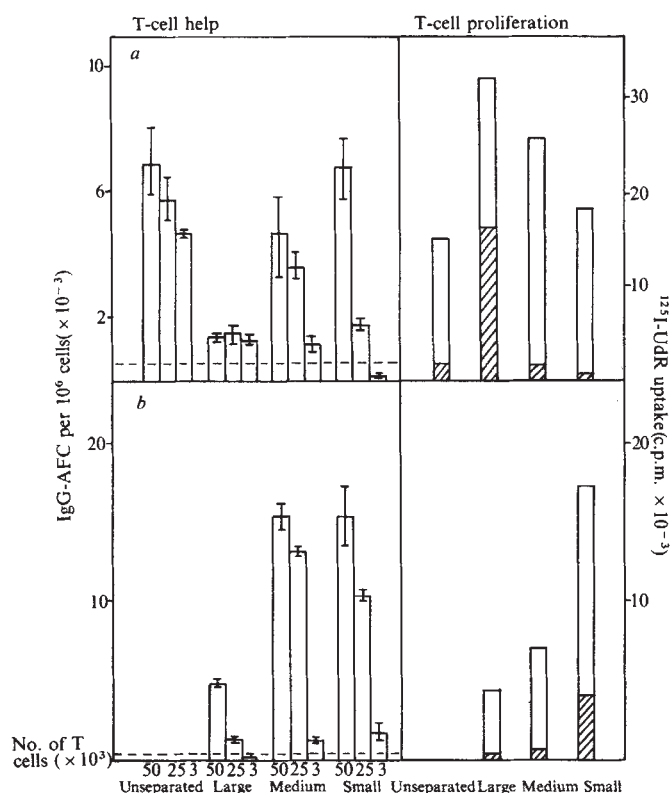


Fig. 1 The expression of helper activity is not dependent on the proliferative responsiveness of T cells. Nylon wool purified T cells (1×10^6) from KLH-primed CBA mice were cultured with starch-induced PEC (5×10^4) and KLH ($1 \mu\text{g ml}^{-1}$) in serum-free medium²⁰ in microtitre plates. After 7 days the cells were collected, counted and 1×10^5 T cells were distributed into the wells of microtitre plates and stimulated with PEC (5×10^3) and KLH ($1 \mu\text{g ml}^{-1}$). More specific details of the production of T-lymphocyte lines are given elsewhere (manuscript submitted). Five days later the cultures were collected and the cells separated according to size by 1g sedimentation on bovine serum albumin gradients²¹ into a large, medium and small fraction (a). To test for helper activity, graded numbers of unseparated and separated cells were added to 4×10^5 anti-Thy 1+C' treated DNP-CGG primed CBA spleen cells (as source of B cells) and incubated with TNP-KLH ($0.02 \mu\text{g ml}^{-1}$) in minimal essential medium containing 0.5% mouse and 1% fetal calf serum. Five days later the IgG anti-DNP (TNP and DNP cross-react at the antibody level) response of two independent cultures was determined and expressed as antibody-forming cells (AFC) per 10^6 'B' cells $\pm$ standard deviation. The dotted lines represent 'B' cells and TNP-KLH alone. Details of the helper assay are given in ref. 22. For the proliferation test 1×10^5 cells of each fraction were incubated with ($\square$) or without ($\blacksquare$) PEC (5×10^3) and KLH ($1 \mu\text{g ml}^{-1}$) for 3 days in serum-free medium. Eight hours before the end of the experiment ^{125}I -2-deoxyuridine (^{125}I -UdR) ($0.25 \mu\text{Ci}$ per well) was added. The results are given as c.p.m. of three independent cultures. Standard deviations were less than 15% and are not included. For all proliferation assays done the dose of antigen as well as the number of accessory cells added were titrated, but for brevity only one parameter of each is shown. The cells of each fraction were independently cultured with PEC and KLH for another 2 weeks and tested again for help and proliferation in the same way as described above (b). During the culture period the unseparated cells became contaminated and could not be tested further. Three more experiments of this type have been performed using other mouse strains (BALB/c and BALB/k) and results very similar to those reported here were obtained.

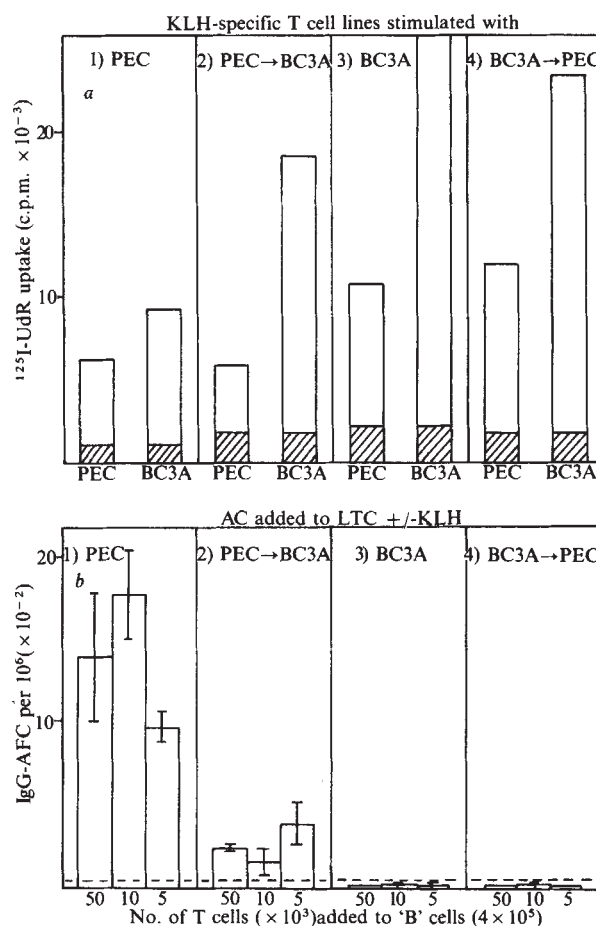


Fig. 2 Accessory cells select the T-cell function to be expressed. KLH-specific T-cell lines from (CBA $\times$ BALB/c) F_1 mice were set up with PEC as described in Fig. 1 legend or with mitomycin-C treated BC3A cells. The BC3A cell line is a B/T-L virus induced, Ia.8 positive leukaemia line of BALB/c origin²³. After 1 week, the cultures were divided and one half was further stimulated with KLH and the same accessory cells as before, the other half with KLH and the opposite accessory cells giving four groups of T-cell lines; (1) PEC stimulated, (2) first PEC and then BC3A stimulated, (3) BC3A stimulated and (4) first BC3A and then PEC stimulated. Two weeks later the cells of each group were tested for proliferation (a) and helper activity (b). For proliferation, 1×10^5 T cells of each group were incubated with ($\square$) or without ($\blacksquare$) PEC or BC3A cells (5×10^3) and KLH ($40 \mu\text{g ml}^{-1}$) for 3 days in serum-free medium. The conditions of ^{125}I -UdR pulsing and the calculations of the data are the same as in Fig. 1. Antigen as well as accessory cell dose-response curves have been done. The helper assay (b) was done as described in Fig. 1 legend using the T-cell lines resulting from the four sequences of stimulation. Similar experiments have been performed using BALB/c or A/J mice as a source of T cells as well as other tumour lines beside BC3A cells, TA3 and D1B, with identical results.

Fig. 3 KLH-pulsed BC3A tumour cells induce proliferation but do not activate helper cells. 1×10^6 BALB/c PEC or mitomycin-C treated BC3A tumour cells were incubated with $100 \mu\text{g}$ KLH for 1 h at 37°C , then washed three times with excess medium to remove unbound KLH. The KLH-pulsed PEC or BC3A cells were tested as accessory cells for induction of antigen-specific proliferation or help using KLH-specific BALB/c T cells. The procedures of priming as well as the purification of T cells are given elsewhere^{19,22}. For proliferation, graded numbers of KLH-pulsed cells as well as non-pulsed cells were added to 3×10^5 T cells and incubated with or without KLH ($40 \mu\text{g ml}^{-1}$) for 3 days. The proliferation assay was done as described in Fig. 1 legend. For helper cell activation pulsed and non-pulsed cells were added to 8×10^5 T cells and incubated with or without KLH ($2 \mu\text{g ml}^{-1}$). After 4 days the cultures were collected, the cells washed and graded numbers were added to 4×10^5 DNP-primed 'B' cells and TNP-KLH as described in Fig. 1 legend. Only the dose yielding the highest AFC numbers is given (2.5×10^4 T cells). The antigen-pulsed BC3A or PEC cells were titrated over a wide range (10^2 – 10^5), but only the most significant doses are given. The standard deviations for the proliferation assay were below 15% for each observation. Identical pulsing experiments were performed with other tumour cell lines, for example, TA3, D1B, P388D1 with similar results.

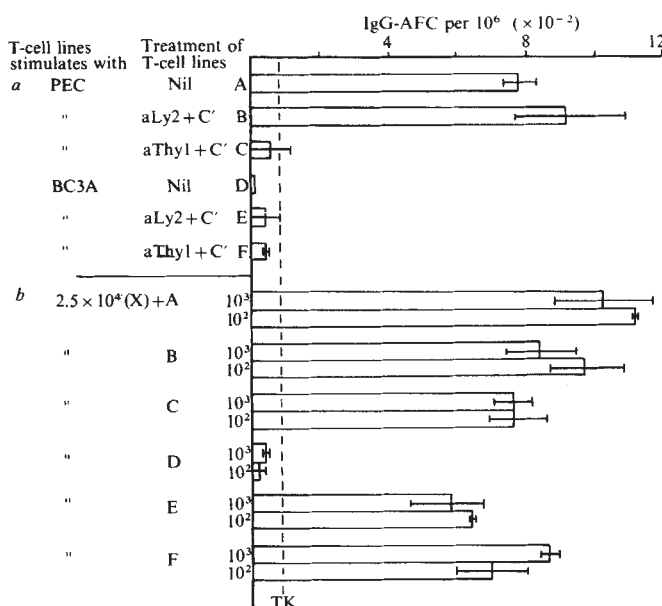
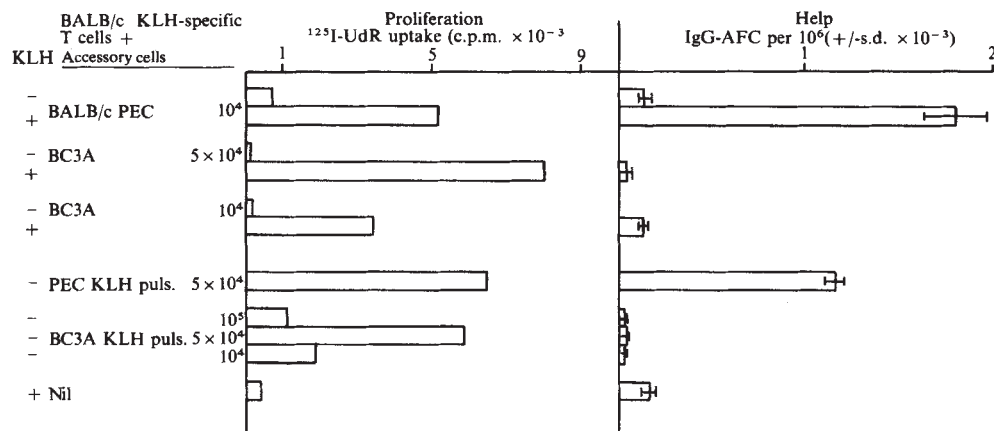


Fig. 4 BC3A cells activate suppressor T cells. KLH-specific T-cell lines from BALB/c mice were set up with PEC or with mitomycin-C treated BC3A cells as described in Fig. 2. Two weeks later the T-cell lines were tested for helper activity as described in Fig. 1 legend. Aliquots of the cells were first treated with monoclonal anti-Thy1 + C' or with monoclonal anti-Ly2 + C' (gifts from Dr M. Feldmann). The treatment with anti-Thy1 + C' resulted in 100% death of the cells as determined by the trypan blue dye exclusion test. Treatment with anti-Ly2 + C' resulted in 30% dead cells of the PEC stimulated T-cell lines and 75% dead cells of the BC3A cell-stimulated T-cell lines. To test for helper activity the cells of all groups were titrated, but only one dose (2.5×10^4) is shown here (a). To test for suppressor activity (b) graded numbers of each group (A to F) were added to a positive control group consisting of 2.5×10^4 PEC-stimulated T-cell lines (X), 4×10^5 'B' cells and TNP-KLH ($0.02 \mu\text{g ml}^{-1}$). Two representative doses (10^3 , 10^2) are shown. The dotted line represents 'B' cells and TNP-KLH alone. Similar experiments have been performed using TA3 or IC-21 tumour cells instead of BC3A cells with identical results.

anti-Ly 2 and complement (C') (Fig. 4a) although they proliferated in response to antigen and accessory cells (data not shown). Similar results were obtained when *in vivo* KLH-primed and *in vitro* restimulated T cells were used instead of T-cell lines.

To test whether suppressor cells were induced, BC3A cell-stimulated T cells were added to PEC-stimulated T-cell lines, B cells and trinitrophenyl (TNP)-KLH. Figure 4b shows that BC3A cell-stimulated T cells strongly suppressed the helper activity, even if added in a very low dose (10^2). If the BC3A cell-stimulated T-cell lines were first treated with anti-Thy 1 + C' their suppressive activity was abolished, demonstrating that the suppressor cell must be a T cell. Also, if the BC3A cell-stimulated T-cell lines were treated with anti-Ly 2 + C' the suppressive activity was markedly diminished. However, the BC3A cell-stimulated T-cell lines also induced suppression of the anti-dinitrophenyl (DNP) response if added into cultures of DNP-CGG primed spleen cells and DNP-CGG, demonstrating that the suppression is not antigen-specific (data not shown).

Our data demonstrate that BC3A cells function as accessory cells for antigen-specific T-cell proliferation, but do not induce antigen-specific T-helper cells; however, they do activate non-specific Ly 2⁺ suppressor T cells which inhibit antibody responses. These functional characteristics are not restricted to BC3A tumour cells, since other tumour lines tested, for example TA3, IC-21, behaved in a very similar manner to the BC3A cells.

We recognize that in studying accessory-T-cell interactions, care should be taken in drawing general conclusions from observations made by measuring a single accessory cell dependent T-cell function. However, our results demonstrate that first, selection of a particular function occurs at the time of *in vitro* T-cell activation and persists, and second, the selection of the function is made by the type of accessory cell available. In addition, we observed that some tumour cell lines are efficient in stimulating T-suppressor cells. Why some accessory cells activate helper cells and others stimulate suppressor cells remains unknown.

We thank Drs M. Feldmann, M. Raff and N. A. Mitchison, University College London for helpful suggestions and for critical reading of the manuscript and Ms S. Studer and S. Mischler for technical help. This investigation was supported by Swiss NSF grant 3.112-0.81.

Received 14 February; accepted 27 May 1983.

1. Rosenthal, A. S. & Shevach, E. M. *J. exp. Med.* **138**, 1194–1212 (1973).
2. Erb, P. & Feldmann, M. *J. exp. Med.* **142**, 460–472 (1975).
3. Miller, J. F. A. P., Vadas, M. A., Whitelaw, A. & Gamble, J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5095–5098 (1975).

4. Erb, P., Stern, A. C. & Cecka, M. J. *Adv. exp. Med.* **155**, 579–590 (1982).
5. Rosenthal, A. S. & Shevach, E. M. *Contemp. Topics Immunobiol.* **5**, 47–90 (1976).
6. Stern, A. C., Erb, P. & Gislis, R. H. *J. Immun.* **123**, 612–615 (1979).
7. Steinman, R. M. & Nussenzeig, M. C. *Immun. Rev.* **53**, 127–147 (1980).
8. Sunshine, G. H., Katz, D. R. & Czitrom, A. A. *Eur. J. Immun.* **12**, 9–15 (1982).
9. Chesnut, R. & Grey, H. J. *Immun.* **126**, 1075–1079 (1981).
10. McKean, D. J. *et al. J. exp. Med.* **154**, 1419–1431 (1981).
11. Cohen, D. A. & Kaplan, A. M. *J. exp. Med.* **154**, 1881–1898 (1981).
12. Glimcher, L. H., Kim, K. J., Green, I. & Paul, W. E. *J. exp. Med.* **155**, 445–459 (1982).
13. Glimcher, L. H. *et al. Nature* **298**, 283–284 (1982).
14. Walker, E. B., Lanier, L. L. & Warner, N. L. *J. Immun.* **128**, 852–859 (1982).
15. Walker, E. B., Lanier, L. L. & Warner, N. L. *J. exp. Med.* **155**, 629–634 (1982).
16. Walker, E. B., Warner, N. L., Chesnut, R., Kappler, J. & Marrack, P. *J. Immun.* **128**, 2164–2169 (1982).
17. Birmingham, J. R. *et al. J. Immun.* **128**, 1491–1492 (1982).
18. Ramila, G., Sklenar, I. & Erb, P. in *Ir Genes: Past, Present and Future*, 445–449 (Humana, New Jersey, 1983).
19. Ramila, G., Studer, S., Mischler, S. & Erb, P. *J. Immun.* (submitted).
20. Schreier, M. H. & Tees, R. *Int. Archs Allergy appl. Immun.* **61**, 227–237 (1980).
21. Anderson, J., Schreier, M. H. & Melchers, F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1612–1616 (1980).
22. Vogt, P., Erb, P., Keller, B. & Feldmann, M. *Eur. J. Immun.* **11**, 377–381 (1981).
23. Chesebro, B., Wehrly, K., Chesebro, K. & Portis, J. *J. Immun.* **117**, 1267–1274 (1976).

Polymorphism in mitogenic effect of IgG1 monoclonal antibodies against T3 antigen on human T cells

W. J. M. Tax, H. W. Willems*, P. P. M. Reekers†, P. J. A. Capel & R. A. P. Koene

Department of Medicine, Division of Nephrology and Departments of Pathology* and Transplantation Serology†, St Radboudziekenhuis, University of Nijmegen, Nijmegen, The Netherlands

It has recently been described that monoclonal antibody OKT 3, that reacts with the T3 determinant on all peripheral T lymphocytes in man, is also mitogenic for T cells^{1,2}. We have produced two monoclonal antibodies (WT 31 and WT 32) which apparently react with the T3 antigen³. We have now tested the mitogenic effect of these antibodies and compared it with the mitogenicity of three other anti-T3 monoclonal antibodies, OKT 3, UCHT 1 (ref. 4) and anti-Leu 4 (ref. 5). Two distinct patterns were observed. WT 32 and OKT 3, both of the IgG2a subclass, were mitogenic for human T cells in all cases studied. By contrast, WT 31, UCHT 1 and anti-Leu 4, all of the IgG1 subclass, were devoid of mitogenic effect in 30% of the individuals tested (non-responders). The mitogenicity of all five anti-T3 antibodies was fully dependent on the presence of monocytes. Addition of purified monocytes from a responder to purified lymphocytes from a non-responder induced responsiveness to both IgG1 and IgG2a anti-T3 antibodies. These results suggest that the polymorphism in the mitogenic effect of these IgG1 antibodies is caused by polymorphism in monocyte function, possibly at the level of the Fc receptor that reacts with mouse IgG1.

When mononuclear cells from six healthy unrelated volunteers were tested for the mitogenic effect of the monoclonal antibodies OKT 3, anti-Leu 4, UCHT 1, WT 31, and WT 32, we found that WT 32 and OKT 3 were mitogenic for the lymphocytes from all six subjects. WT 31, anti-Leu 4, and UCHT1, however, induced no mitogenic response in two individuals (Fig. 1), although the binding of these antibodies to lymphocytes was similar (both in percentage of positive cells and in fluorescence intensity) in all six subjects (data not shown). The 'non-responsiveness' towards antibodies WT 31, anti-Leu 4, and UCHT 1 was reproducible with different blood samples and with frozen cells from the same donor. Furthermore, the same patterns of reactivity were found when we used two other batches of fetal calf serum, or human serum, although the amount of thymidine incorporation was influenced by the serum source.

We subsequently tested 54 normal subjects, 16 (30%) of whom appeared to be non-responders with respect to the mitogenic effect of WT 31 (Table 1). This group of 'non-

responders' also gave a poor response to UCHT 1. The proliferative response of these individuals to WT 32 and OKT 3 was always at least nine times higher than the response to WT 31 or UCHT 1. In a responder, the stimulation indices for WT 31 and UCHT 1 were always higher than 5, and the responses to each of the antibodies WT 31, WT 32, OKT 3 and UCHT 1 were almost the same. In all cases tested (10 individuals from the responder group and 5 from the non-responder group), the mitogenic effect of anti-Leu 4 (another IgG1 monoclonal antibody against the T3 antigen) was concordant with the mitogenic effect of WT 31 and UCHT 1. Therefore, we conclude that antibodies WT 32 and OKT 3 are always mitogenic, but that some individuals exhibit no mitogenic response to WT 31, UCHT 1 or anti-Leu 4. As antibodies WT 31, UCHT 1 and anti-Leu 4 all have immunoglobulin subclass IgG1 whereas WT 32 and OKT 3 are IgG2a, the non-responsiveness appears to be related to the Fc moiety of IgG1.

As both the IgG1 and the IgG2a antibodies bind to the same antigen, one might expect that preincubation of mononuclear cells from a non-responder with one of the IgG1 anti-T3 antibodies inhibits the proliferative response to the IgG2a antibodies. This proved to be the case. As shown in Table 2, preincubation of non-responder cells with UCHT 1 or anti-Leu 4 prevented the response to WT 32 or OKT 3. Preincubation with WT 31 was also inhibitory in this respect, but inhibition was not complete. This is consistent with our previous data which showed that WT 31 can inhibit, but not completely block, the binding of radiolabelled WT 32 to T cells³. Mononuclear cells from a responder still showed a proliferative response after preincubation with IgG1 antibody, as would be expected since both IgG1 and IgG2a antibodies are mitogenic for these cells.

To investigate a possible involvement of monocytes, we purified lymphocytes and monocytes from responders and non-responders, and then measured the proliferative response of various cell mixtures. Mononuclear cells from peripheral blood were separated into lymphocytes and monocytes by counterflow centrifugation (elutriation) monitored by continuous flow-cytometry⁶. As shown in Table 3, the proliferative response of purified lymphocytes in the absence of monocytes was low. Addition of purified autologous monocytes to responder lymphocytes induced responsiveness to all four monoclonal antibodies (expt 1). Importantly, addition of non-responder monocytes induced a response to WT 32 and OKT 3 only, but not to the IgG1 antibodies WT 31 and UCHT 1. When lymphocytes from a non-responder were tested (expt 2), the addition of autologous monocytes only induced a proliferative response to WT 32 or OKT 3. Responder monocytes, however, were necessary to establish responsiveness to WT 31 or UCHT 1. The interpretation of these data is complicated by the simultaneous occurrence of incipient mixed lymphocyte reactions due to the presence of allogeneic monocytes. Although the

Table 1 Non-responsiveness to mitogenic effect of antibodies WT 31 and UCHT 1

		Responders (n = 38)	Non-responders (n = 16)
SI	WT 31	16 (5–57)	1.3 (0.6–2.2)
	UCHT 1	15 (5–46)	1.6 (0.9–2.5)
	WT 32	17 (4–43)	20 (6–44)
	OKT 3	17 (6–48)	21 (7–44)
Ratio: WT31/WT 32		0.9 (0.5–1.4)	0.02 (0.0–0.10)
UCHT 1/OKT 3		0.9 (0.6–1.6)	0.04 (0.0–0.12)

Mean values are given, with ranges between parentheses. Stimulation indices (SI) were calculated by dividing c.p.m. of incorporated ³H-thymidine (measured as described in Fig. 1 legend) obtained in the presence of monoclonal antibody, by background c.p.m. obtained in the absence of antibody. Net c.p.m. incorporated in the presence of WT 31, were divided by net c.p.m. obtained with WT 32, giving the ratio WT 31:WT 32 for each individual. Similar calculations were performed to give the ratio UCHT 1: OKT 3.

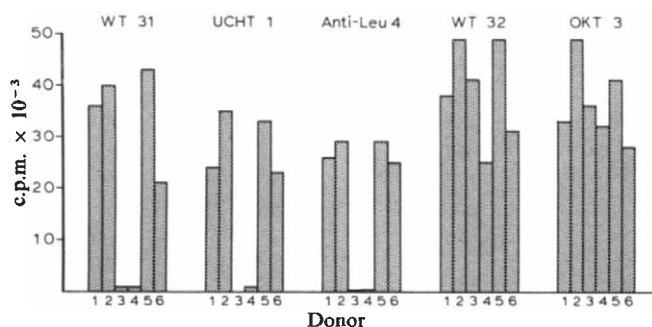


Fig. 1 Mitogenicity of monoclonal antibodies against T3 antigen. The mitogenic effect of monoclonal antibodies OKT 3 (Ortho Diagnostics), anti-Leu 4 (Becton-Dickinson), UCHT 1, WT 31 and WT 32 was tested by incubating an optimal concentration of antibody with 10^5 mononuclear cells in U-bottomed wells of microtitre plates (Costar). Total volume was 150 μ l, and the medium was RPMI 1640 with 10% fetal calf serum containing 2 mM glutamine, 1 mM sodium pyruvate and 50 μ g ml⁻¹ gentamicin. Incubation was for 72 h at 37 °C and 5% CO₂. 0.5 μ Ci ³H-thymidine (Amersham, specific activity 26 Ci mmol⁻¹) was added 18 h before the end of the incubation. The incorporated radioactivity was collected on glass fibre filters using a cell collector (Titertek) and counted in a liquid scintillation counter (Packard). All determinations were performed in triplicate. Net c.p.m. were calculated by subtracting the incorporation measured in the absence of monoclonal antibody from the total incorporation.

Table 2 Preincubation of non-responder cells with IgG1 antibody inhibits proliferative response to IgG2a antibody

Cells	Preincubation	Incubation	Mitogenic response (net c.p.m.)	Inhibition
NR	—	WT 32	34,668	—
	WT 31	WT 32	12,472	64%
	UCHT 1	WT 32	4,502	87%
	Anti-Leu 4	WT 32	5,026	85%
NR	—	OKT 3	35,991	—
	WT 31	OKT 3	13,472	63%
	UCHT 1	OKT 3	4,766	87%
R	—	OKT 3	38,192	—
	WT 31	—	24,498	—
	WT 31	OKT 3	31,978	—

Cells from a non-responder (NR) or from a responder (R) were incubated with IgG1 antibody for 15 min at room temperature. WT 32 or OKT 3 was added and the proliferative response was measured as described in Fig. 1 legend.

results had been corrected for thymidine incorporation in the absence of monoclonal antibody, we felt it necessary to include an additional control. Therefore, monocytes from a non-responder (B2) were added to lymphocytes from another non-responder (B1). Although this combination gave rise to a similar mixed lymphocyte reaction to the combination of non-responder lymphocytes with responder monocytes (data not shown), the corrected values given in Table 3 (expt 2) indicate that this cell mixture remained unresponsive to WT 31 or UCHT 1. Therefore, we conclude that the polymorphism found in the mitogenic effect of WT 31 and UCHT 1, is related to polymorphism of monocyte function. One possible explanation could be that monocytes of non-responder type eliminate lymphocytes coated with WT 31 or UCHT 1, by antibody-dependent cell-mediated cytotoxicity. When, however, non-responder monocytes were added to a mixture of lymphocytes and monocytes from a responder, the response to WT 31 and UCHT 1 was not decreased (Table 3, expt 3). Similarly, a mixture of lymphocytes from a non-responder and responder monocytes showed a proliferative response to WT 31 and UCHT 1 that was not affected by the addition of monocytes from the non-responder (expt 4). These findings strongly argue against the occurrence of lymphocyte killing by non-responder monocytes.

We next investigated whether the polymorphism in responsiveness to the IgG1 antibodies was linked to some other polymorphic trait. Population studies revealed no correlation with age, sex, ABO blood group or one of the HLA-A, B or DR antigens. When the responsiveness to WT 31 and UCHT 1 was measured in recipients of kidney grafts from living related donors, we found that the family donors had the same responder status as the respective recipients in 11 out of 12 cases. This is highly suggestive of a hereditary character of the phenomenon. Family donors had been selected on the basis of three criteria: HLA-identity, MLC non-reactivity and compatibility with respect to ABO blood group. Family studies are in progress to investigate whether the trait 'non-responsiveness to WT 31' segregates with one of these characteristics.

In conclusion, we found that IgG1 monoclonal antibodies against the T3 antigen do not induce mitosis in mononuclear cell suspensions from about 30% of normal subjects, although these antibodies are potent mitogens for other individuals. To our knowledge, this is the first report of non-responsiveness to a mitogenic compound despite binding of this compound. All five antibodies are reactive with the same cell-surface antigen^{3,7,8}, but differ in immunoglobulin subclass. The non-responsiveness to IgG1 antibodies appears to be related to the Fc moiety of these antibodies, and reveals a new type of polymorphism in monocyte function. This polymorphism presumably involves an Fc receptor. Further investigations will be

Table 3 Role of monocytes on mitogenic effect of anti-T3 antibodies

Cell mixtures				Mitogenic effect (net c.p.m.)			
Expt no	Lymphocytes	Monocytes	Monocytes	WT 31	UCHT 1	WT 32	OKT 3
1	A	—	—	1,087	<0	6,199	824
	A	A	—	36,881	38,033	35,472	43,371
	A	B1	—	214	<0	19,197	21,950
2	B1	—	—	474	85	3,221	376
	B1	B1	—	<0	<0	22,734	19,259
	B1	B2	—	3,442	<0	18,384	16,769
	B1	A	—	16,311	21,353	20,563	24,182
3	A	A	—	36,881	38,033	35,472	43,371
	A	A	B1	38,603	35,562	31,487	40,310
	A	A	B2	35,520	36,550	31,771	41,840
4	B1	A	—	16,311	21,353	20,563	24,182
	B1	A	B1	20,779	22,840	20,961	29,055

A is a responder to all anti-T3 antibodies. B1 and B2 are non-responders to WT 31, UCHT 1 and anti-Leu 4. Mononuclear cells were isolated from peripheral blood using Ficoll gradient centrifugation, and then separated into lymphocytes and monocytes by counterflow centrifugation⁶ (elutriation). The number of lymphocytes was 7×10^5 per well, and, when indicated, 10^4 monocytes were added. The proliferative response was measured and net c.p.m. were calculated as described in Fig. 1 legend.

necessary to elucidate the precise mechanism responsible for the observed non-responsiveness to IgG1 antibodies. Furthermore, it is of interest to study the possible biological or clinical significance of this phenomenon, because polymorphism in monocyte function might have important consequences for immune responsiveness.

This work was supported by the Dutch Kidney Foundation. We thank Dr P. Beverley for the gift of monoclonal antibody UCHT 1, Dr A. Hoitsma for performing the FACS analysis, and co-workers of the Department of Haematology for assistance in the counterflow centrifugation.

Received 28 March; accepted 31 May 1983.

1. Wauwe, J. P. van, My, J. R. de & Goossens J. G. *J. Immun.* **124**, 2708-2713 (1980).
2. Chang, T. W., Kung, P. C., Gingras, S. P. & Goldstein, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1805-1808 (1981).
3. Tax, W. J. M., Leeuwenberg, H. F. M., Willems, H. M., Capel, P. J. A. & Koene, R. A. P. in *Human Leucocyte Markers Detected by Monoclonal Antibodies* (Springer, Berlin, in the press).
4. Beverley, P. C. L. & Callard, R. E. *Eur. J. Immun.* **11**, 329-334 (1981).
5. Ledbetter, J. A. et al. *J. exp. Med.* **153**, 310-323 (1981).
6. De Mulder, P. H. M. et al. *J. immun. Meth.* **47**, 31-38 (1981).
7. Burns, G. F., Boyd, A. W. & Beverley, P. C. L. *J. Immun.* **129**, 1451-1457 (1982).
8. Borst, J., Prendiville, M. A. & Terhorst, C. *J. Immun.* **128**, 1560-1564 (1982).

A conserved sequence in the immunoglobulin J_{κ} - C_{κ} intron: possible enhancer element

**Laurent Emorine*, Michael Kuehl†, Lawrence Weir‡,
Philip Leder‡ & Edward E. Max*§**

* Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205, USA

† National Cancer Institute, Naval Medical Oncology Branch,
Bethesda, Maryland 20014, USA

‡ Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

Several functionally important genetic elements (such as the TATA box, mRNA splice sequences, poly(A) addition signal) were first detected as short segments of unexplained sequence homology within non-coding regions of different genes. A short region of unknown sequence in the intron between the human J_{κ} and C_{κ} immunoglobulin coding regions was found to be sufficiently homologous to the corresponding segment of the mouse gene to form stable heteroduplexes¹. Although no specific function has yet been definitely ascribed to this region (which we call the kappa intron conserved region, or KICR), some functional significance has been inferred from the findings that (1) activation of B lymphocytes induces a DNase hypersensitivity site in this region^{2,3} and (2) deletions including this region reduce expression of κ genes introduced into lymphoid cells⁴. To delineate the KICR more precisely and to test the generality of the sequence conservation in a third species, we have sequenced this region of the human and mouse genes and have examined the corresponding region of a recently cloned rabbit κ gene. We find a segment of about 130 base pairs (bp) that shows striking conservation in all three species, demonstrating homology significantly higher than within the C_{κ} coding region itself. Two short sequences from the J_{κ} - C_{κ} intron that were noted by other investigators to be homologous to proposed 'enhancer' sequences both lie within the conserved region.

Molecular cloning of the mouse, human and rabbit genomic κ genes has been previously described^{1,5,6}. The sequencing strategy for the present analysis of the $J_{\kappa}-C_{\kappa}$ intervening sequence is shown in Fig. 1. The resulting three sequences were compared pairwise using the dot matrix display described by Maizel and Lenk⁷. A composite of the three matrices aligned for comparison of the homologous regions is shown in Fig. 2. The diagonal strip that bisects the figure was cut from a matrix

in which the mouse sequence (x axis) was compared with the human (y axis). The upper portion of a rabbit-versus-mouse matrix is shown displaced vertically, while a human-versus-mouse matrix is shown displaced horizontally to the left of the central strip. The dots forming 45° diagonal lines demonstrate the homologous regions between the compared species. At about 700 bp 5' to the C region all three matrices show a KICR diagonal which, surprisingly, appears stronger than the diagonal for the corresponding C region coding block. This unexpected finding is confirmed by the quantitative comparisons shown in Table 1, for which the boundaries of the KICR have been arbitrarily defined (see Fig. 3) as a minimal segment showing high homology in all three species. These boundaries need not coincide exactly with boundaries of a conserved functional element since (1) homology may extend fortuitously somewhat beyond the region actively conserved and (2) the function of the element may tolerate some departure from strict sequence conservation.

To determine whether KICR-related DNA segments are present in other immunoglobulin genes we probed Southern blots of cloned human immunoglobulin μ^8 and λ^9 genes with a human KICR probe (320 bp *PvuII*-*PstI* fragment; see Fig. 1). No cross-hybridization was detected even under low washing stringency (room temperature, $0.1 \times \text{SCC}$) and long autoradiographic exposures (data not shown). Furthermore, although KICR probes from each species detected the known κ gene bands (from all three species) in Southern blots of total genomic DNA, no other bands were visualized.

What function might such a highly conserved region have in the intervening sequence between J_K and C_K ? We have considered the possibility that the KICR might encode a protein analogous to the 'mRNA maturase' gene found within the intervening sequence of the yeast mitochondrial *box* (cytochrome *b*) gene¹⁰. However, examination of the sequence in the vicinity of the KICR for open reading frames in either orientation revealed none longer than 114 bp that is conserved between the three species. If the KICR does not encode a polypeptide it might nevertheless be independently transcribed to produce an RNA species analogous to 'identifier sequences'

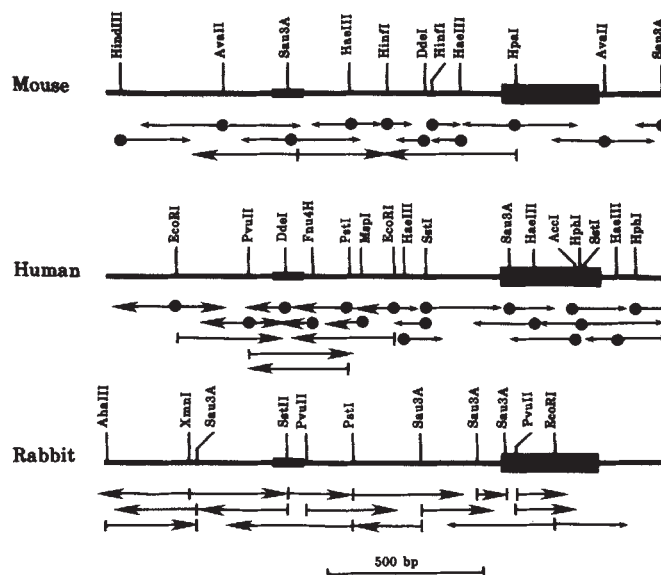


Fig. 1 Sequencing strategy. Previous sequence data on the mouse⁵, human¹ and rabbit⁶. Germ-lines genes were extended for the present analysis. The three bold lines indicate the three genes, with the KICR and C_K coding regions shown as bold rectangles. The arrows indicate individual sequence runs. Short vertical bars represent the beginning of M13 subclones sequenced by the dideoxy terminators method^{21,22}, while the black circles represent the ³²P-labelled ends of fragments sequenced by the method of Maxam and Gilbert²³. Small arrowheads indicate sequences previously determined, while the large arrowheads represent sequences determined for the present paper.

§ To whom correspondence should be addressed.

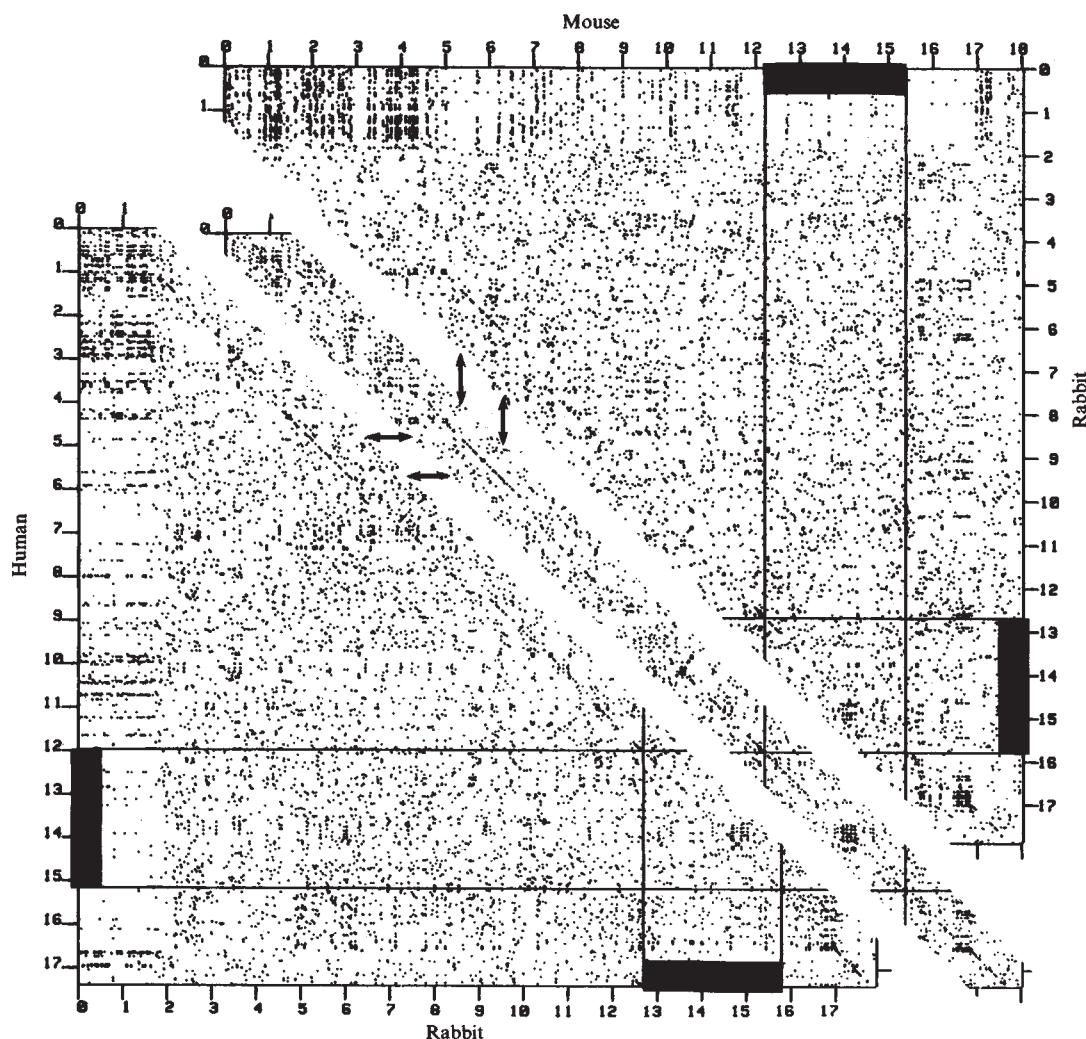


Fig. 2 Matrix comparisons. The three pairwise comparisons between the mouse, human and rabbit sequences are displayed using the matrix method of Maizel and Lenk⁷. Each dot represents a stretch of four nucleotides identical between the sequences represented on the horizontal and vertical axes. The central diagonal strip represents a comparison between mouse (x axis) and human (y axis). The black rectangles on the axes represent the C_{κ} coding region. Each numbered division on the axes represents 100 bp. Double arrows indicate a region that is conserved in all three species. The matrix also detects an additional short region of lesser homology further 3' at about position 800 in the human sequence, 830 (mouse) and 850 (rabbit).

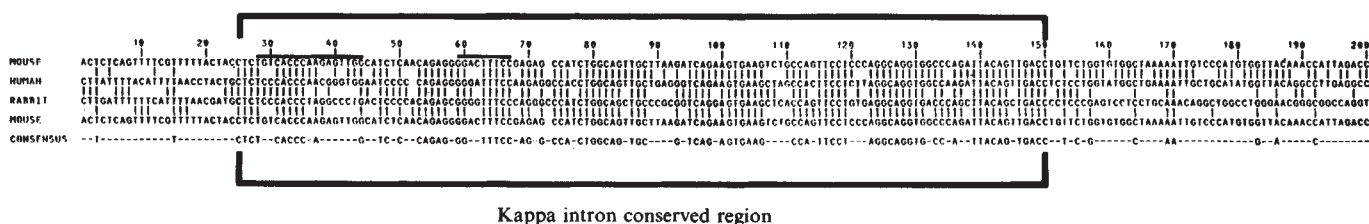


Fig. 3 Sequence comparisons. The corresponding regions of the J_{κ} - C_{κ} intervening sequence of mouse, human and rabbit are shown; vertical bars designate identical nucleotides. The sequence showing high homology in all three species has been designated kappa intron conserved region (KICR) and corresponds to the region indicated by double arrows in Fig. 2. The borders of the KICR are somewhat arbitrary; the human-mouse homology appears to extend somewhat further 3' and all three species show additional upstream homology if multiple gaps are inserted. To maximize homology within KICR a single space was inserted in the mouse sequence at a position that remains uncertain and at position 52 in the human sequence. The two bars above the mouse sequence highlight short DNA segments similar to regions within putative enhancer sequences as described in the text.

described by Sutcliffe *et al.*¹¹ or the 'activator sequences' postulated by Britten and Davidson¹². Northern blot analysis of total RNA from four human lymphoid cell lines using the human KICR probe failed to reveal such an independent transcript. No hybridization was seen to RNA from cell lines 8402 (T cell), REH (B-cell precursor) or CLL-DC (μ , λ B cell). In RNA from BHM-23 (μ , κ B cell) a faint doublet was observed at about 4 kilobases (kb) which co-migrated with a region of faint hybridization detected with a C_{κ} coding probe and which was assumed to represent incompletely processed κ transcript (data not shown). It is possible that RNA from other cells or tissues might contain the KICR sequences as independent transcripts, although roles for the KICR as either 'identifier sequence' or

Table 1 Sequence conservation is higher in the KICR than in the κ constant region

	KICR	C_{κ}
Rabbit-human	77%*	66%†
Human-mouse	81%	69%
Mouse-rabbit	72%	60%

* Per cent nucleotide identity in the 126 bp KICR segment indicated in Fig. 2.

† Per cent nucleotide identity within the published C_{κ} coding region sequences mouse⁵, human¹ and rabbit⁶.

'activator sequence' seem to be weakened by absence of detectable homology to other segments in the genome on Southern blots.

A conserved sequence that is neither translated nor independently transcribed could still have regulatory roles for which precedents have been described, either (1) as part of a κ RNA transcript—for example as an attenuator-like element, or by forming secondary structure to mediate splicing¹³—or (2) at the DNA level—for example as a promoter, operator or enhancer sequence.

Enhancer elements were first recognized in viruses such as simian virus 40 (SV40) and Moloney sarcoma virus, where they occur in tandem duplicated sequences of about 72 bp and can stimulate transcription from a variety of positions in the viral genome and in both orientations; the second duplicate copy of these sequences is not required for their enhancing effect^{14,15}. A role for the KICR as an enhancer sequence seems particularly attractive as it might explain the dramatic stimulation of *V* region gene transcription that occurs when a particular *V* recombines with the *J-C* locus¹⁶. In fact, evidence is now accumulating for such an enhancer sequence in this region of the κ gene of mouse (ref. 4 and S. L. Morrison and V. T. Oi, personal communication) and man (L.W., P.L. and H. Potter, unpublished results) and in the homologous region of the murine μ gene^{17,18}. The sequence of the implicated DNA segment in the *J_H- μ* intron shows minimal overall homology to the KICR sequence, as expected from our failure to detect any cross-hybridization between the human KICR and a human μ clone.

The exact sequences necessary for enhancer function have not been established for any of the elements now recognized. Two short segments in the mouse *J_κ-C_κ* intron have been noted to bear homology to important regions of proposed enhancers, and our sequence comparisons indicate that both of these segments fall within the region conserved between the three species. (See segments highlighted by bars above and the mouse sequence in Fig. 3.) The first segment bears homology to the polyoma sequence TcTcACCCAgGccTaG¹⁹ (where capital letters indicate identity with the mouse KICR sequence nucleotides 28 to 44). This polyoma sequence falls within a 137 bp *PvuII* fragment which appears functionally homologous to the SV40 enhancer region (although minimal sequence homology is apparent between the two regions). A single point mutation converting the TaG to TGG is found in mutants with extended host range, which possibly results from improved enhancer function. The second segment indicated in Fig. 3 is homologous (in reverse orientation) to a segment within the 72 bp SV40 enhancer (TGTGGAAAGTCC) which includes the 'core' sequence TGG ^{TTT}AAA G suggested by Laimins *et al.*²⁰ as a possible

consensus element within various enhancer regions. This 'core' sequence also appears within the immunoglobulin heavy-chain segment thought to contain an enhancer¹⁷.

Whether any of the short homologies discussed above reflect minimal sequence requirements for enhancer function is unclear at present. Certainly these short segments cannot account for the conservation of the much longer KICR sequence between human, mouse and rabbit. Furthermore the significance of the DNase hypersensitivity which occurs in approximately this region of the *J_κ-C_κ* intron remains unknown. Plasmid constructions with various alterations in this region of the κ gene are currently being investigated in several laboratories; these should help to resolve the role of the conserved region and should eventually allow the replacement of the noncommittal term KICR with an appropriate functional name.

We thank Jacob Maizel and John Owens for their help with the computer analysis of sequence homologies, Virginia Shaw for editorial assistance, and Vernon Oi, Elwood Linney, Susumu Tonegawa, Robert Deans, Randolph Wall and George Khoury for helpful discussions and permission to cite unpublished results. This work was supported in part by grants from the ABC Research Foundation and E. I. DuPont de Nemours Inc.

Received 8 June; accepted 21 June 1983.

- Hieter, P. A., Max, E. E., Seidman, J. G., Maizel, V. J. Jr & Leder, P. *Cell* **22**, 197–207 (1980).
- Parslow, T. G. & Granner, D. K. *Nature* **299**, 449–451 (1982).
- Chung, S.-Y., Folson, V. & Wooley, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2427–2431 (1983).
- Queen, C. & Baltimore, D. *Cell* (in the press).
- Max, E. E., Maizel, J. V. Jr & Leder, P. *J. biol. Chem.* **256**, 5116–5120 (1981).
- Emorine, L., Dreher, K., Kindt, T. J. & Max, E. E. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Maizel, J. V. Jr & Lenk, R. P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7665–7669 (1981).
- Ravetch, J. V., Siebenlist, U., Korsmeyer, S., Waldmann, T. & Leder, P. *Cell* **27**, 583–591 (1981).
- Hieter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. *Nature* **294**, 536–540 (1981).
- De La Salle, H., Jacq, C. & Slonimski, P. P. *Cell* **28**, 721–732 (1982).
- Sutcliffe, J. G., Milner, R. J., Bloom, F. E. & Lerner, R. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4942–4946 (1982).
- Britten, R. J. & Davidson, E. H. *Science* **165**, 349–357 (1969).
- Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. *Nature* **300**, 719–724 (1982).
- Gruss, P., Dhar, R. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 943 (1981).
- Benoist, C. & Chambon, P. *Nature* **290**, 304 (1981).
- Mather, E. L. & Perry, R. P. *Nucleic Acid Res.* **9**, 6855 (1981).
- Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* (in the press).
- Banerji, J., Olson, L. & Schaffner, W. *Cell* (in the press).
- Laimins, L. A., Kessel, M., Rosenthal, N. & Khoury, G. in *Enhancers and Controlling Elements* (ed. Gluzman, Y.) (Cold Spring Harbor Laboratory, New York, 1983).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Messing, J. & Vieira, J. *Gene* **19**, 269–276 (1982).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499 (1980).

The pyrogenic and mitogenic actions of interleukin-1 are related

Gordon W. Duff & Scott K. Durum

Departments of Internal Medicine and Pathology and the Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut, 06510 USA

Our present understanding of the pathogenesis of fever is that host macrophages, following activation by an appropriate stimulus^{1,2} such as Gram-negative lipopolysaccharide (LPS)³ immune complexes⁴, or primed lymphocytes in the presence of specific antigen⁵, synthesize and release endogenous pyrogen (EP). EP is carried in the blood circulation to the hypothalamic area of the brain where its action, involving a protein synthetic step⁶, results in an increase of the level at which body temperature is maintained⁷. Recently, it was shown^{8,9} that EP is very similar and possibly identical to another macrophage mediator previously called lymphocyte activating factor and now known as interleukin-1 (IL-1)¹⁰ which, in conjunction with lectin¹¹ or specific antigen¹², induces clonal expansion of T lymphocytes. We show here that murine T-cell proliferation in response to IL-1 *in vitro* is greatly increased when the cells are exposed to a temperature typical of fever and that injection of the same IL-1 causes fever in mice. If this relationship exists *in vivo*, the resulting facilitation of a T-cell-dependent immune response may well confer survival value and contribute to the evolutionary conservation of fever—a phylogenetically ancient response to infection.

Purified IL-1 was prepared from LPS-stimulated murine macrophage-like tumour cells (P388D1)¹³. After 4 days, supernatants were precipitated with 65% saturated ammonium sulphate, dialysed against phosphate-buffered saline and fractionated by gel filtration chromatography over G75 superfine Sephadex. Interleukin-2 (IL-2) was generated by culture of the T-cell hybridoma FS6-14.13 (10⁶ cells per ml)¹⁴ or rat spleen cells (10⁷ per ml) with concanavalin A (Con A), 2 µg ml⁻¹, for 24 h without serum. Supernatants were checked for IL-2 activity by maintenance of either Con A-induced blast cells¹⁵ or the IL-2-dependent T-cell line HT-2 (ref. 16). T-cell proliferation assays were performed using thymocytes and spleen cells from C3H/HeJ mice (LPS unresponsive) or spleen cells from C57BL/6J (LPS responsive) mice. T-cell cultures in 200 µl RPMI with 10% fetal bovine serum, Con A (1 µg ml⁻¹) and different concentrations of IL-1 or IL-2 were kept at 37 or

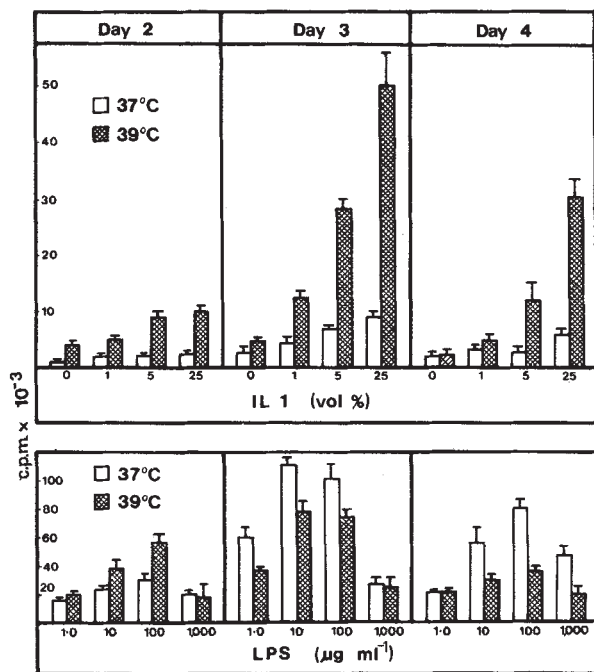


Fig. 1 The proliferative responses of C3H/HeJ thymocytes (1×10^6 in 200 μ l) to different concentrations of purified murine IL-1 and Con A ($1 \mu\text{g ml}^{-1}$ in all cultures) (upper panel), and the response of C57BL/6J splenic B cells (5×10^4 in 200 μ l) to different concentrations of LPS (*Escherichia coli*; Difco) (bottom panel). Cultures were maintained for 2, 3 or 4 days at either 37 or 39 °C (as indicated), then pulsed with $1 \mu\text{Ci } ^3\text{H}$ -thymidine for 5 h at 37 °C. Means $\pm$ s.e. of radiolabelled thymidine uptake for triplicate cultures are shown.

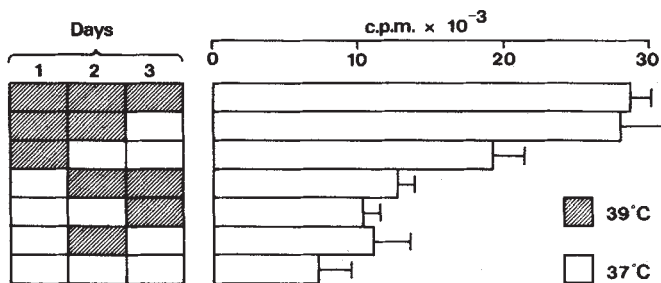


Fig. 2 The proliferative responses of C3H/HeJ thymocytes (1×10^6 in 200 μ l) to purified murine IL-1 (5% final volume) and Con A ($1 \mu\text{g ml}^{-1}$). Cultures were maintained at 37 or 39 °C for 3 days or were moved between 37 and 39 °C after the first or second days of culture as shown on the left of the figure. After 3 days, proliferation was assayed as before (means $\pm$ s.e. of triplicate cultures are given).

Table 1 Hyperthermia increases T-cell proliferation induced by IL-1 or IL-2

Mitogen Type	Vol %	Thymus		Spleen	
		37 °C	39 °C	37 °C	39 °C
IL-1	0	5.3 (0.6)*	24.7 (2.2)	5.0 (0.5)	22.4 (2.5)
	1	37.0 (1.8)	154.0 (33)	7.7 (0.4)	35.5 (2.8)
	5	83.8 (4.9)	226.0 (9.4)	18.2 (2.7)	62.5 (8.5)
	25	48.0 (4.8)	214.0 (36)	13.2 (0.7)	55.2 (3.1)
IL-2	0	2.2 (0.7)	20.0 (3.1)	8.9 (0.3)	32.0 (7.7)
	1	3.2 (0.3)	19.0 (2.9)	28.0 (2.9)	106.0 (5.4)
	5	11.7 (0.2)	108.0 (3.4)	34.0 (5.7)	186.0 (12.0)
	25	41.0 (1.5)	102.0 (6.1)	51.6 (2.9)	203.0 (4.3)

All cultures contained Con A at $1 \mu\text{g ml}^{-1}$. Thymocytes from C3H/HeJ mice, 1×10^6 per 200 μ l culture, were used. Spleen cells were from C57BL/6J in the IL-1 experiment, C3H/HeJ in the IL-2 experiment, both at 5×10^4 per 200 μ l culture.

* Counts per minute $\times 10^{-3}$; means $\pm$ s.e. (in parentheses) of triplicate cultures.

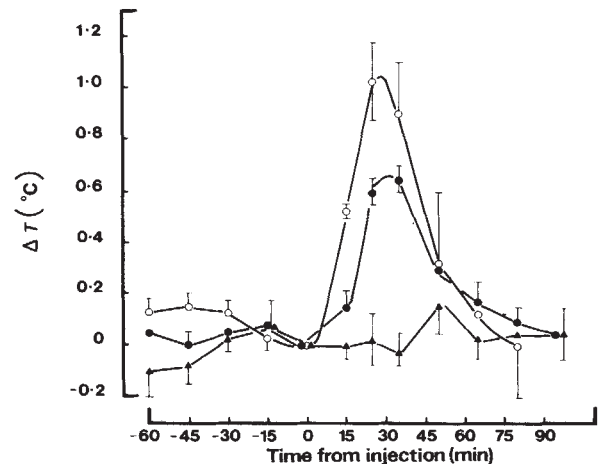


Fig. 3 The pyrogenic response of C3H/HeJ mice to purified murine IL-1 (intravenous) at two concentrations: $\circ$, 100%; $\bullet$, 10%. Control injections were culture medium alone ($\blacktriangle$). Mice were kept at an ambient temperature of 33 °C throughout. The mean ($\pm$ s.e.) changes in rectal temperatures (ΔT) from those at the time of injection (time 0) for a group of four mice are plotted against time from injection.

39 °C (± 0.3) in 5% CO_2 incubators for 3 days unless indicated otherwise in the figures. They were then pulsed for 5 h with $1 \mu\text{Ci } ^3\text{H}$ -thymidine (at 37 °C in all cases) and proliferation was assayed as incorporated radioactive thymidine. Splenic B-cell proliferative responses to the B-cell mitogen LPS were tested in the same way (for details see figure legends).

To assay the pyrogenicity of IL-1, groups of C3H/HeJ mice were caged individually in an incubator set at 33 °C. Rectal temperatures were recorded by thermistor at 15-min intervals. When the mouse's temperature had been stable for 1 h, IL-1 was injected into the tail vein and temperature recordings were continued at 10- then 15-min intervals for 90 min¹⁷. Fever curves were drawn as the mean difference in temperature from that at the time of injection.

We first found that thymocyte and splenic T-cell proliferation in response to purified IL-1 and Con A was markedly increased at 39 °C (Table 1, upper panel). The increase was usually 4–5-fold but in some experiments with thymocytes 10-fold increases were seen.

If the T-cell mitogenic effect of IL-1 is indirect, and requires IL-2 production¹⁵, the effect of hyperthermia on IL-1-induced T-cell proliferation may represent increased IL-2 production at the higher temperature or an increased T-cell proliferative response to IL-2 or both. We attempted to distinguish between these possibilities by testing the effect of temperature on IL-2-induced T-cell proliferation and on IL-2 production in cultures of mouse spleen cells with Con A ($2 \mu\text{g ml}^{-1}$). As seen in Table 1 (lower panel), both thymic and splenic T-cell proliferation induced by IL-2 from two different sources was increased at 39 °C. As with the response to IL-1, the thymocyte augmentation was up to 10-fold while spleen cell proliferation increased 4 to 5-fold. The production of IL-2 from 24-h Con A/spleen cell cultures (10^6 – 10^7 cells per culture) was assayed by maintenance of the IL-2-dependent T-cell line HT2. The difference in IL-2 production from cultures performed at 39 °C compared with 37 °C was never more than 1.5–2-fold (data not shown). This experiment, however, only measured excess IL-2 available for transfer in culture supernatants. It is possible, and even likely, that IL-2 consumption by the spleen cell culture itself was increased and that this increased IL-2 turnover was not reflected in supernatant IL-2 concentration. Experiments are in progress to clarify this. We can surmise, therefore, that the hyperthermic enhancement of IL-1 induced T-cell proliferation was probably accounted for by an increased response to IL-2 but cannot exclude an additional contribution from significantly increased IL-2 production at the higher temperature.

To determine whether the temperature dependence of T-cell proliferation was a property of lymphocytes in general, we examined splenic B-cell proliferation in response to the B-cell mitogen LPS. Figure 1 shows results from a characteristic experiment comparing IL-1-induced T-cell proliferation and LPS-induced B-cell proliferation. The T-cell response to IL-1 and Con A (Fig. 1, top panel) peaks on the third day of culture and shows significant hyperthermic augmentation that is present by day 2 and maintained to day 4. Thus, hyperthermia magnifies the T-cell response and does not merely accelerate it.

With LPS-induced B-cell proliferation (Fig. 1, bottom panel), again peak responses occurred on day 3. A small (less than twofold) thermal enhancement was usual in 2-day cultures but at both days 3 and 4, culture at 39 °C reduced the proliferative response, indicating that the increased proliferation of T cells cultured at 39 °C is not a property of all lymphocytes.

For maximal thermal enhancement of 3-day T-cell cultures, it was not necessary to maintain the hyperthermia throughout the culture period (Fig. 2). Exposure to hyperthermia on only the first day of culture gave substantially increased proliferation and cultures held at 39 °C for the first 2 days did not benefit by a third day of hyperthermia. Conversely, if exposure to 39 °C did not occur during the first day of culture, no significant thermally increased proliferation occurred even when the cells were held at 39 °C for days 2 and 3 of culture. The early temperature-sensitive event(s) may result in increased numbers of activated T cells that can respond to IL-1 or IL-2.

Finally, we determined the pyrogenicity of the same IL-1 that was used in the T-cell proliferation assays in mice of the same strain as those used as a source of T cells in these assays. The dose-dependent fever curves produced in a group of four C3H/HeJ mice following intravenous injection of IL-1 or vehicle alone (RPMI) are shown in Fig. 3. A characteristic short-latency, transitory fever was produced. We also tested two strains of LPS-sensitive mice, (C57BL/6 $\times$ DBA/2) F_1 and C57BL/6J, and obtained similar fevers. In all, 40 mice of three different strains were tested in a euthermic environment and a mean ($\pm$ s.e.) basal rectal temperature of 37.28 ± 0.1 °C was recorded. Of these, 16 received IL-1 and developed a mean peak fever of 38.59 ± 0.2 °C (range 0.4–1.9 °C).

Our results show that purified, murine tumour-produced IL-1 was both a potent pyrogen and an immunostimulant in mice. Moreover, we found that these two actions of IL-1 may be related, in that the *in vitro* co-mitogenic action of IL-1 was greatly enhanced at hyperthermic temperatures similar to those generated *in vivo* by the pyrogenic action of IL-1. Hyperthermia resulted in an increase in the magnitude of the T-cell proliferative response to IL-1 and Con A and not just an acceleration. Temperature-sensitive events occurred early in the course of T-cell culture and the effect could be explained, at least in part, by increased proliferation in response to IL-2, although an additional augmentation of IL-1-induced IL-2 production at the higher temperature was not excluded.

The role of fever in host defence has been the subject of several recent reviews^{1,2,18}. The presence of fever *in vivo* was correlated with survival in experimentally infected lizards¹⁹, fish²⁰ and rabbits²¹ and accelerated the inflammatory response at the infection site²². The beneficial effects of hyperthermia on lymphocyte proliferation in response to lectins or antigens *in vitro* have been described previously (reviewed in ref. 18) and some of our early findings have been summarized elsewhere^{23,24}. Similar findings to ours regarding the temperature dependence of IL-1-induced mitogenesis have been independently reported recently²⁵. It seems that the host mediator, IL-1, by virtue of its T-cell mitogenic activity, is an important cofactor in the generation of an immune response and the pyrogenic action of this same mediator on the brain ensures optimal thermal conditions for T-cell clonal expansion.

Depending on the subsets of T cells involved, thermal enhancement of IL-1-dependent mitogenesis could have far-reaching functional consequences. For example, we have found that the primary immune response of murine spleen cells to

sheep erythrocytes *in vitro* was greatly increased at 39 °C^{23,26}. This was attributable to a beneficial effect of hyperthermia on the development of Ly 1 helper T cells and did not involve reduced generation of suppressor cells or any change in the response to pre-formed helper or suppressor factors. Exposure of T cells to hyperthermia for as little as 15 min caused marked changes in protein synthesis, with the appearance of four new prominent protein bands²⁷. We are currently attempting to correlate the appearance of these proteins with the temperature-induced changes in antibody production that we have described.

In conclusion, if the relationship we report between temperature and IL-1-induced T-cell mitogenesis *in vitro*, applies to immune responses *in vivo*, this effect may contribute significantly to the survival value of fever. In specific instances of human disease, fever may be desirable or not—depending on whether the role of the immune response is in host defence or in the pathogenesis of the disease.

We thank Drs E. Atkins and R. K. Gershon for their help and support throughout this project. The HT2 cell line was kindly given by Dr J. Ryan and the FS6-14.13 by Drs J. Kappler and P. Marrack. We thank Ms JoAnn Somers for secretarial assistance. This work was supported by NIH grants A101564, CA29606 and CA14216.

Received 30 March; accepted 19 May 1983.

- Atkins, E. *Yale J. Biol. Med.* **55**, 283–289 (1982).
- Dinareello, C. A. in *Advances in Host Defense Mechanisms* Vol. 1 (eds Gallin, J. I. & Fauci, A. S.) 57–75 (Raven, New York, 1982).
- Duff, G. W. & Atkins, E. *J. Immun. Meth.* **52**, 323–331 (1982).
- Duff, G. W., Gekowski, K. & Atkins, E. *Clin. Res.* **30**, 364A (1982).
- Atkins, E., Feldman, J. D., Francis, L. & Hursh, E. *J. exp. Med.* **135**, 1113–1132 (1972).
- Cranston, W. I., Hellon, R. F. & Townsend, Y. *J. Physiol., Lond.* **305**, 337–344 (1980).
- Cranston, W. I., Duff, G. W., Hellon, R. F. & Mitchell, D. *J. Physiol., Lond.* **257**, 767–777 (1976).
- Rosenwasser, L. J., Dinarello, C. A. & Rosenthal, A. S. *J. exp. Med.* **150**, 709–714 (1979).
- Murphy, P. A., Simon, P. L. & Willoughby, W. F. *J. Immun.* **124**, 2498–2501 (1980).
- Oppenheim, J. J. & Gery, I. *Immun. Today* **3**, 113–119 (1982).
- Gery, I., Gershon, R. K. & Waksman, B. H. *J. Immun.* **107**, 1778–1785 (1971).
- Durum, S. K. & Gershon, R. K. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4747–4750 (1982).
- Lachman, L. B., Blyden, A. T., Hacker, M. & Handschumacher, R. E. *Cell. Immun.* **34**, 416–419 (1976).
- Harwell, L., Skidmore, B., Marrack, P. & Kappler, J. *J. exp. Med.* **152**, 893–904 (1980).
- Larsson, E. L., Iscove, N. N. & Coutinho, A. *Nature* **283**, 664–666 (1980).
- Leibson, H. J., Marrack, P. & Kappler, J. W. *J. exp. Med.* **154**, 1681–1693 (1981).
- Bodel, P. & Miller, H. *Proc. Soc. exp. Biol. Med.* **151**, 93–96 (1976).
- Roberts, N. J. *Microbiol. Rev.* **43**, 241–259 (1979).
- Kluger, M. J., Ringler, D. H. & Anver, M. R. *Science* **188**, 166–168 (1975).
- Covert, J. B. & Reynolds, W. W. *Nature* **267**, 43–45 (1977).
- Kluger, M. J. & Vaughn, L. K. *J. Physiol., Lond.* **282**, 243–251 (1978).
- Bernheim, H. A., Bodell, P. T., Askew, P. W. & Atkins, E. *Br. J. exp. Path.* **59**, 76–84 (1978).
- Duff, G. W. & Durum, S. K. *Clin. Res.* **30**, 694A (1982).
- Duff, G. W. & Durum, S. K. *Yale J. Biol. Med.* **55**, 437–442 (1982).
- Hanson, D. F., Murphy, P. A., Silicano, R. & Shin, H. S. *J. Immun.* **130**, 216–221 (1983).
- Jampel, H. D., Duff, G. W., Gershon, R. K., Atkins, E. & Durum, S. K. *J. exp. Med.* **157**, 1229–1238 (1983).
- Dzierzak, E. A., Duff, G. W., Janeway, C. A., Bottomly, K. & Durum, S. K. *Clin. Res.* **30**, 694A (1982).

Somatic inactivation of genes on chromosome 13 is a common event in retinoblastoma

R. Godbout*, T. P. Dryja†, J. Squire†, B. L. Gallie†§ & R. A. Phillips*†§

Departments of * Medical Biophysics and § Ophthalmology, University of Toronto, † The Ontario Cancer Institute, Toronto, Canada M4X 1K9

‡ Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts 02114, USA

Through family studies¹ and analysis of patients with congenital chromosome abnormalities², the germ-line mutation responsible for the hereditary form of the eye tumour, retinoblastoma, has been assigned to the q14 region on chromosome 13 and closely linked to an enzyme called esterase D (ESD). Knudson has proposed that as few as one somatic event in addition to

§ To whom correspondence should be addressed.

the germ-line mutation is required to induce tumours in patients with the hereditary form of retinoblastoma³; the non-hereditary form requires two somatic events to occur in the same cell. The somatic event(s) may involve either mutation of the remaining normal gene at 13q14 or mutation of a gene at another site in the genome. Here we have examined six retinoblastoma patients who are heterozygous for electrophoretic variants of ESD. Although the normal cells of all six patients expressed both variants, the tumour cells of four patients expressed enzyme from only one of the two ESD alleles. We tentatively conclude that induction of a retinoblastoma tumour requires the somatic inactivation of genes near the ESD locus including the remaining normal gene at the retinoblastoma (RB) locus.

Karyotypic analysis of more than 30 retinoblastoma tumours both in our laboratory⁴ and by Kuznetsova *et al.*⁵ has failed to show a high frequency of abnormalities of chromosome 13; in only one case was there a somatic deletion involving the 13q14 locus. However, Balaban *et al.*⁶ found that five of six retinoblastomas showed deletion at the 13q14 locus. Benedict *et al.*⁷ have reported a retinoblastoma that lacks one chromosome 13. The patient has a possible congenital deletion, undetectable by karyotypic analysis but resulting in half-normal quantitative levels of ESD in normal cells. The chromosome 13 missing from the tumour carried the remaining normal genes at the q14 locus and the tumour had no ESD activity. Therefore, they proposed that the second event involves the deletion of the remaining functional RB locus.

It is possible that small alterations occur on chromosome 13 that are undetectable by even the most sensitive chromosome banding techniques. As ESD is known to be closely linked to the RB locus, we studied ESD expression in tumour cells to search for possible somatic changes at this locus. Because tumour heterogeneity makes quantitative measurements of ESD levels difficult, we looked for electrophoretic variants of ESD in normal cells and tumour cells obtained from retinoblastoma patients who were heterozygous for ESD variants. If cells of a heterozygote fail to express one of the two alleles, the electrophoretic pattern will reflect this loss of function.

By screening peripheral blood or fibroblasts from the patients whose tumours were available for investigation, we identified six individuals whose normal cells contained both electrophoretic variants, ESD-1 and ESD-2. ESD is probably a dimer because it shows the typical three-banded pattern expected of a dimeric protein: the top band is 2-2, the middle band 1-2 and the bottom band 1-1 (Fig. 1).

Retinoblastoma cells were obtained directly from surgical specimens or from tissue culture (Table 1). Two tumours were analysed directly from the patients, one from a surgically removed eye (HL163) and the other from a metastatic brain lesion obtained at autopsy (RB430). Tumours RB409 and RB355 had been grown in the eyes of nude mice before growth in tissue culture⁸. RB405-A and -B and RB412 were placed on fibroblast feeder layers immediately after surgery⁹ and were subsequently grown without feeder cells. RB405-A and -B were established from separate pieces of the same tumour and maintained as separate cultures. To avoid contamination with normal cells, only tumours that had become independent of their feeder layer were used for the ESD studies. All the tumours which could be analysed have unique chromosome abnormalities which allow them to be distinguished from one another. No karyotype data were obtained from HL163.

All six tumour samples tested had ESD activity. Two tumours, RB405 (A and B) and RB355 (both unilateral tumours and probably non-heritable), retained the electrophoretic pattern characteristic of heterozygous cells (Table 1, Fig. 1). In contrast, the other four tumours failed to express enzyme activity from one of the ESD alleles. RB409 (bilateral, therefore hereditary) lost ESD-1 activity, while RB412, RB430 (both unilateral) and HL163 (bilateral) lost ESD-2 activity.

Analysis of the chromosomes in five tumours using the Giemsa-trypsin banding technique¹⁰ revealed no deletions involving chromosome 13 (Fig. 2). High-resolution banding^{11,12}

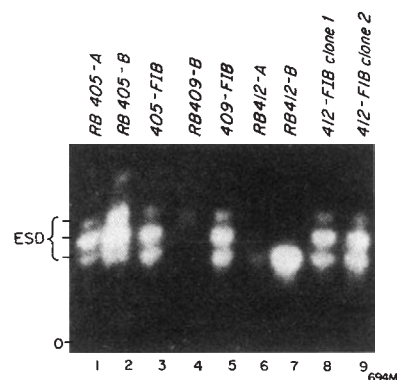


Fig. 1 Starch gel electrophoresis was carried out as described by Hopkinson *et al.*¹⁹ using phosphate buffer, pH 6.5, and staining with methylumbelliferyl acetate. Lanes 1, 2, 4, 6 and 7 represent ESD patterns of RB tumour cells. Both RB409 and RB412-A and B have lost activity from one allele and are therefore lacking two of the three bands expected of heterozygous cells. Lanes 3, 5, 8 and 9 represent ESD patterns in the fibroblasts of the corresponding patients with RB.

of RB412, which lacks ESD-2 activity, also failed to reveal deletion in the 13q14 region. However, we emphasize that small sub-band deletions in tumours may not be detected using present cytogenetic methods. Although one chromosome 13 may appear slightly shorter in an occasional spread, we attribute this anomaly to uneven contraction rather than deletion. We have only considered deletion to be present when the majority of spreads showed the same abnormality.

Some of the tumours studied had been grown in tissue culture or in nude mice for up to 1½ yr before analysis for ESD activity. As there must be continuous selection for fast growing cells, it is possible that loss of ESD activity occurred during tissue culture. To evaluate this possibility, cells from two tumours frozen immediately after enucleation were thawed, grown to mass culture and tested. The ESD patterns for both RB409-B and RB412-B were identical to those obtained from the long-term cultures (Table 1). In addition, two tumours which did not express ESD-2, RB430 and HL163, were analysed as fresh tumour specimens immediately after removal from the patient. Therefore, it is unlikely that random loss of ESD activity during long-term cultures explains the above results; the loss probably occurred before removal of the tumour from the patient.

Three different processes could account for the somatic loss of ESD activity: (1) The tumour cell may have a somatic mutation, deletion or chromosome rearrangement involving the

Table 1 Expression of ESD in retinoblastoma cells

Tumour	Patient type	Source of tumour	ESD pattern in tumour
RB409-A	Bilateral	A → B → C	2
RB409-B	Bilateral	B → C	2
RB412-A	Unilateral	A → B → C	1
RB412-B	Unilateral	B → C	1
RB405-A	Unilateral	B → C	1-2
RB405-B	Unilateral	B → C	1-2
RB355	Unilateral	A → C	1-2
RB355-clone 7	Unilateral	A → C	1-2
RB430	Unilateral	D	1
HL163	Bilateral	D	1

Tumours were studied from four sources: A, grown in the eyes of nude mice; B, grown in tissue culture with a feeder layer; C, grown in tissue culture without a feeder layer; D, obtained directly from the patient. The total elapsed time between removal of the tumour and establishment of a free-growing culture ranged from 3 months to 2 yr. Enzyme activity was measured only in tumours obtained from sources C and D.

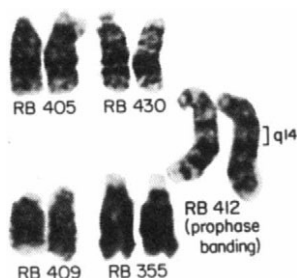


Fig. 2 Partial G-banded karyotypes of five RB tumours used in this study showed both chromosome 13s without apparent deletion. Methotrexate-synchronized¹¹ high-resolution prophase banding (550 band level)¹² of chromosomes from RB412 failed to demonstrate any deletion at 13q14 (bracket).

ESD locus. (2) The individuals involved may be mosaics for a 13q14 deletion, with the tumour cells being derived from cells hemizygous for one of the ESD alleles; Motegi has presented evidence for mosaicism in several retinoblastoma patients¹³. (3) During evolution of the tumour, the cells may have lost one chromosome 13 and then duplicated the remaining chromosome, making both chromosomes identical at every locus. We have obtained preliminary evidence against the last two possibilities. To study the possibility of mosaicism, seven separate clones of fibroblasts from patient 412 were isolated; all expressed both ESD alleles (two clones are shown in Fig. 1). In addition, 10 fibroblast clones from patient 409 expressed both ESD alleles. Obviously many more clones must be examined before we can conclude that the patients are not mosaic. The possibility of chromosome loss and duplication was studied by analysis of naturally occurring heteromorphisms on chromosome 13. In many individuals, including our patient 412, the short arm of one of the two chromosome 13s has a fluorescent satellite, allowing the two chromosomes to be distinguished. RB412 probably contained both of the original chromosomes because one chromosome 13 contained a fluorescent satellite and the other did not. The same pattern was observed in the fibroblasts of this patient.

The above results allow us to draw three conclusions concerning genetic changes associated with retinoblastoma. First, it is likely that the ESD and RB loci are distinct but closely linked. Although Howard speculated that ESD may be identical to the RB gene¹⁴, we found two tumours that expressed both ESD alleles with no evidence of alteration in enzyme activity. Furthermore, one of these tumours (RB355) was cloned, and the subclone, RB355-7, retained both ESD activities, eliminating the possibility that the original tumour sample contained two populations of cells, one expressing ESD-1 and the other ESD-2.

Second, the resultant loss of ESD activity in the tumour cells must reflect the occurrence of somatic genetic events. The normal cells from all individuals tested expressed both alleles, indicating that the loss of ESD activity was not a germ-line change. Given the known close proximity of the ESD and RB loci, it is possible that the genetic events leading to loss of ESD activity are related to the proposed somatic events required to initiate tumour formation. In this context, it is important to know whether or not the somatic event has occurred on the chromosome carrying the RB gene or on that carrying the remaining normal homologous allele. If this change represents the second somatic event leading to retinoblastoma tumour, one would predict that it would occur on the normal chromosome and result in inactivation of the remaining normal gene at the RB locus. Alternatively, the original germ-line mutation at the RB locus may create a chromosomal instability that increases the probability of genetic changes in linked genes such as ESD. Unfortunately, it is not possible to test which chromosome is altered with the patients presently available for study. Identification of the ESD variant linked to the RB gene

will require families with both affected and unaffected individuals in more than one generation and with an informative polymorphism for ESD. In our study, none of the patients had a previous family history of retinoblastoma, and the bilaterally affected patients, presumed to carry a germ-line mutation, were probably new mutations.

Third, the observation of somatic loss of ESD activity in both unilateral (probably non-heritable) and bilateral (heritable) retinoblastoma implies that the same genetic changes are responsible for both non-heritable and heritable tumours. Balaban *et al.*⁶ previously reported somatic loss of 13q14 in both unilateral and bilateral patients by karyotypic analysis. Our data on ESD support their conclusion that the second event in retinoblastoma involves somatic inactivation of genes at the 13q14 locus and that the same genetic changes occur in both the heritable and non-heritable forms of the disease. A recent study of the expression of polymorphic enzymes (including ESD) in many non-retinoblastoma tumours suggests that heterozygosity was detected less frequently in tumours than would be expected on the basis of the known heterozygosity in the population¹⁵. Although these investigators concluded that somatic inactivation of autosomal genes frequently occurs in tumours, they presented no direct comparisons of enzyme activity in normal and tumour tissue from the same individual. The present study, however, provides the first clear evidence of somatic inactivation of a gene linked to a known human cancer-causing gene.

Knudson has presented several lines of evidence³, drawn from clinical data, that induction of retinoblastoma requires two genetic changes. Analysis of patients carrying various chromosome abnormalities indicated that the germ-line mutation represents loss of gene function at the RB locus. Comings¹⁶ has proposed that autosomal dominant hereditary tumours such as retinoblastoma represent the inheritance of a defective regulatory or suppressor gene and that tumours develop when mutations inactivate the normal suppressor gene. Benedict *et al.*⁷ have proposed a similar model. Our data support the conclusion of Comings in that loss of ESD activity may represent a somatic genetic change leading to loss of gene expression in the region of the chromosome carrying the normal RB gene which is probably a regulatory gene. Mutation of regulatory genes may allow subsequent overexpression of a putative oncogene, but it is not necessary to postulate that the overexpressed oncogene carries any mutation. Thus, retinoblastoma may represent a cancer caused by defective gene regulation rather than production of a dominant mutant oncogene as has been proposed for human bladder cancer^{17,18}.

This work was supported by grants from the NCI of Canada, Ontario Cancer Treatment and Research Foundation and The Hospital for Sick Children Foundation.

Received 9 March; accepted 7 June 1983.

1. Sparkes, R. S. *et al. Science* **208**, 1042-1044 (1980).
2. Sparkes, R. S. *et al. Science* **219**, 971-973 (1983).
3. Knudson, A. G. Jr *Semin. Oncol.* **5**, 57-60 (1978).
4. Gardner, H. A., Gallie, B. L., Knight, L. A. & Phillips, R. A. *Cancer Genet. Cytogenet.* **6**, 201-211 (1982).
5. Kusnetsova, L. E., Prigogina, E. L., Pogozian, H. E. & Belkina, B. M. *Hum. Genet.* **61**, 201-204 (1982).
6. Balaban, G., Gilbert, F., Nichols, W., Meadows, A. T. & Shields, J. *Cancer Genet. Cytogenet.* **6**, 213-221 (1982).
7. Benedict, W. F. *et al. Science* **219**, 973-975 (1983).
8. Gallie, B. L., Albert, D. M., Wong, J. J. Y., Buyukmichi, N. & Puliafito, C. A. *Invest. Ophthalmol.* **16**, 256-259 (1977).
9. Gallie, B. L., Holmes, W. & Phillips, R. A. *Cancer Res.* **42**, 301-305 (1982).
10. Worton, R. G. & Duff, C. *Meth. Enzym.* **58**, 322-344 (1979).
11. Yunis, J. J., Sawyer, J. R. & Ball, D. W. *Chromosoma* **67**, 293-307 (1978).
12. Yunis, J. J. *Hum. Path.* **12**, 494-503 (1981).
13. Motegi, T. *Hum. Genet.* **61**, 95-97 (1982).
14. Howard, R. O. *Lancet* **ii**, 490-491 (1982).
15. Dracopoli, N. C. & Fogh, J. J. *Natn. Cancer Inst.* **70**, 83-87 (1983).
16. Comings, D. E. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3324-3328 (1973).
17. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. *Nature* **300**, 149-152 (1982).
18. Tabin, C. J. *et al. Nature* **300**, 143-148 (1982).
19. Hopkinson, D. A., Mestriner, M. A., Cortner, J. & Harris, H. *Ann. hum. Genet.* **37**, 119-137 (1973).

Tissue and cell type-specific expression of two human *c-onc* genes

Rolf Müller*‡, Joanne M. Tremblay*,
Eileen D. Adamson† & Inder M. Verma*

* Molecular Biology and Virology Laboratory, The Salk Institute,
PO Box 85800, San Diego, California 92186, USA

† Cancer Research Center, La Jolla Cancer Research Foundation,
10901 North Torrey Pines Road, La Jolla, California 92037, USA

Cellular genes containing nucleotide sequences homologous to retroviral oncogenes (*v-onc*) have been identified in the genomes of a variety of species of both the vertebrate and invertebrate phyla¹. Expression of such genes, termed *c-onc* genes, has been shown to be tissue-specific in chickens¹⁻⁴ and mice⁵⁻⁸ and to be modulated during murine development⁹⁻¹⁰ and differentiation of human haematopoietic cells^{9,10}. We report here that the level of the *c-fos* gene transcripts is 100-fold greater in human term fetal membranes than in other normal human tissues and cells. These levels of *c-fos* expression in human amniotic and chorionic cells are close to the level of *v-fos* expression that results in the induction of osteosarcomas in mice and transformation of fibroblasts *in vitro*. This observation suggests that the induction of neoplastic transformation by the FBJ murine osteosarcoma virus may require the expression of the *fos* gene product at high levels in an inappropriate cell type. In contrast, the human *c-fms* gene is expressed at high levels specifically in term placenta and trophoblastic cells. The tissue and cell type-specific patterns of *c-fos* and *c-fms* expression suggest that the physiological function of the *c-fos* and *c-fms* encoded proteins may be associated with those embryo-derived cells whose primary functions are protection and nourishment of the human fetus.

Our strategy involved the isolation of RNA from human tissues and cell lines and analysis by a quantitative dot blot technique^{5,6} for hybridization to *v-onc*-specific probes. Similar results were generally obtained when both total and poly(A)-containing RNA (poly(A)⁺ RNA) isolated from the same tissue sample were used in parallel experiments. Poly(A)⁺ RNA from selected samples was also analysed by agarose gel electrophoresis followed by Northern blotting to confirm and extend the results obtained by dot blot analysis.

Transcripts from the cellular homologue *c-fos*¹¹⁻¹³ of the oncogene of FBJ murine osteosarcoma virus were found at approximately 10–100-fold higher levels in human term fetal membranes (chorion and amnion) than in other normal tissues (kidney, liver, lymph nodes, lung), in normal fetal fibroblasts (WI-38 cell line)¹⁴ or in the human teratocarcinoma cell line PA-1, which is similar to embryonic ectoderm¹⁵ (Table 1). Similarly, *c-fos* expression was approximately 3–20-fold greater in human term placenta than in other normal tissues or cells (Table 1). Relative to WI-38 and PA-1 cells, the level of *c-fos* transcripts was also elevated in a trophoblastic cell line (BeWo)¹⁶ derived from a gestational choriocarcinoma of the human fetal placenta (Table 1). Expression of *c-fos* in all amniotic cell lines analysed (WISH¹⁷, FL¹⁸, AV-3) was found to be less than 10% of that in human term amniotic tissue (Table 1). This finding may suggest a cell type-specific transcription of *c-fos* within the amnion. Alternatively, the cell lines analysed here could be representative of amniotic cell types at a relatively early stage of differentiation where *c-fos* expression may still be low. In this context it is noteworthy that increasing levels of *c-fos* transcripts have been observed in the developing mouse amnion (R.M., J.M.T., E.D.A. and I.M.V., unpublished observation). High levels of *c-fos* transcripts were also observed in the human vulva carcinoma cell line A431¹⁹. Expression of *c-fos* in A431 cells was determined to be ~50% that observed in human term chorion and amnion (Table 1). It remains uncertain whether the high level of *c-fos* expression in this cell line

is characteristic of the normal cell type from which the vulva carcinoma originated, or whether it occurred during the process of neoplastic transformation or establishment of the cell line. Analysis of a variety of other human tumour cell lines (for example, HT1080²⁰ and HL60²¹) and solid human tumours (lung and kidney carcinoma, lymphosarcoma, fibrosarcoma), however, revealed considerably lower levels of *c-fos* expression than observed in the A431 cell line (Table 1 and data not shown). Note that the high levels of *c-fos* expression in both A431 and amniotic cells may deserve particular attention, as in both cases it seems to be correlated with a high epidermal growth factor binding capacity^{19,22}.

The mRNA of the human *c-fos* gene has a size of approximately 2.2 kilobases (kb, Fig. 1a). This transcript could be identified in human term placenta and chorion as well as in BeWo cells (Fig. 1a, lanes 1, 6, 3). Another *c-fos* transcript of ~3.5 kb was detected in placenta, albeit at a considerably lower concentration than the 2.2-kb RNA (Fig. 1a, lane 1). The size of this transcript would be compatible with that of the unspliced *c-fos* mRNA precursor¹². In the same experimental conditions, *c-fos* transcripts were not detected in human lymph nodes, WI-38 and PA-1 cells (Fig. 1a, lanes 4, 5). To estimate the level of *c-fos* expression in human term fetal membranes, the concentration of *c-fos* transcripts in chorion and amnion was compared with the concentration of *v-fos* transcripts in an osteosarcoma induced in a CBA mouse by FBJ murine osteosarcoma virus (FBJ-MSV) and in an FBJ-MSV-transformed rat cell line (RS2)²³. Table 1 shows that the level of *fos*-related transcripts in the virus-induced tumour and in RS2 cells is less than twice that in chorion and amnion (after correction of the measured values for the approximately twofold stronger hybridization of the *v-fos*-specific probe to *c-fos* (mouse) sequences compared with the human cellular homologue). These results indicate that a high level of the normal cellular *fos* protein is not sufficient to induce neoplasia in cells of human fetal membranes. It remains to be investigated whether amniotic and chorionic cells are able to 'resist' a possibly oncogenic potential of the *c-fos* protein in other cell types, or whether the *fos* gene product displays oncogenic properties only in a structurally aberrant form. Nucleotide sequence analyses have shown that the *v-fos* protein and its cellular counterpart have entirely different carboxy termini, the result of a deletion of 104 nucleotides in the *v-fos* gene^{11,12}. However, the demonstration of similar concentrations of *fos* transcripts in human amnion and chorion and FBJ-MSV-transformed cells indicates that high expression of *c-onc* genes can occur in normal cellular metabolism, and does not necessarily lead to malignant transformation. The induction of neoplasia may require the expression of normal *c-onc* gene product(s) at high levels in an inappropriate cell type or the expression of structurally altered protein(s).

The cellular counterpart *c-fms*²⁴ of the oncogene of the McDonough strain of feline sarcoma virus (SM-FeSV) was also found to be expressed in a tissue and cell type-specific fashion. High levels of *c-fms* transcripts could only be detected in human term placenta and in the trophoblastic BeWo cell line (Table 1). In term fetal membranes the concentration of *c-fms* transcripts was about one-fifth that in placenta (Table 1). The level of *fms* transcripts in placenta and BeWo cells was estimated to be ~20% that observed in NIH/3T3 cells transformed by molecularly cloned SM-FeSV DNA²⁴ (Table 1). Transcripts of ~3.7 kb from the *c-fms* (human) gene could be identified in BeWo cells (Fig. 1b, lane 1), but not in WI-38 cells and normal lymph nodes (Fig. 1b, lanes 2, 3). In contrast to the tissue-specific patterns of *c-fos* and *c-fms* expression, transcripts from the cellular homologues *c-Ki-ras*²⁵, *c-Ha-ras*²⁵ and *c-myc*²⁶ of the oncogenes of Kirsten sarcoma virus, Harvey sarcoma virus and avian myelocytomatosis virus, respectively, were detected at nearly constant levels in all tissues and cells listed in Table 1 (data not shown).

The patterns of *c-fos* and *c-fms* expression suggest specific functions of the proteins encoded by these genes in human

‡ Present address: EMBL, Postfach 10.2209, D-6900 Heidelberg, FRG.

Table 1 Relative levels of *c-onc* transcripts in human tissues and cells compared with virus-transformed cells

Tissue/cell line	Origin	Relative expression of	
		<i>fos</i>	<i>fms</i>
Amnion	Term conceptus	100, 91*	14
Chorion	Term conceptus	80, 100*	22
Placenta	Term conceptus	8, 13, 30, 19†	100
Kidney	Normal adult tissue	≤2	≤10
Liver	Normal adult tissue	≤4	≤10
Lymph nodes	Normal adult tissue	≤4, ≤3*	≤10
Lung	Normal adult tissue	≤3, ≤3*	≤10
BeWo	Choriocarcinoma ¹⁶ (trophoblast-like)	5	80
WISH	Normal amnion ¹⁷	≤2	≤4
FL	Normal amnion ¹⁸	≤2	≤4
AV-3	Normal amnion	5	≤4
WI-38	Fetal fibroblasts ¹⁴	≤1	≤4
PA-1	Teratocarcinoma (ectoderm-like) ¹⁵	≤1	7
A431	Vulva carcinoma ¹⁹	44, 58*	≤4
HT1080	Fibrosarcoma ²⁰	≤1	≤4
Mouse osteosarcoma	Induced by FBJ-MSV	160‡	ND
RS2	FBJ-MSV-transformed rat fibroblasts ²³	150‡	ND
3T3-FMS2	SM-FeSV-transformed NIH/3T3 cells	ND	450‡

Human tissues were homogenized in a buffer consisting of 10 ml of 1% SDS, 0.01 M EDTA, 0.05 M Tris-HCl, pH 8.0, 0.15 M NaCl and 1/20th volume of vanadyl ribonucleoside solution (BRL) per mg of tissue using a motor-driven tissue homogenizer (Virtis) at maximum speed. The homogenate was extracted twice with phenol/chloroform/isoamylalcohol (25:24:1) and once with chloroform/isoamylalcohol (24:1). Nucleic acids were precipitated with 2 vols of ethanol at -20°C overnight. The precipitate was recovered by centrifugation at 10,000 g for 15 min and dissolved in 0.4 ml of water. After addition of 2.6 ml of guanidine hydrochloride buffer⁶ and 0.075 ml of 1 M acetic acid, the RNA was precipitated with 0.5 vols of ethanol at -20°C overnight. After centrifugation the RNA was extracted from the pellet with water, precipitated with ethanol and redissolved in water as described previously⁶. Isolation of RNA from cell lines, selection for poly(A)-containing RNA and quantitative dot blot analyses involving hybridization of RNA to *v-onc* probes were carried out as previously reported⁶. All tissues and virus-transformed cells were analysed as total RNA, all other cell lines as poly(A)⁺ RNA. In some cases both total and poly(A)⁺ RNA from the same sample were analysed in order to correlate different experiments with one another. The relative expression values indicated in the table represent absorbance values which were determined by densitometer scanning of dot blot autoradiograms and normalized to 100 for the highest absorbance value observed in normal tissues or cells in the case of *fos* and *fms*, respectively. In some instances the results from independent experiments using different tissue samples are listed. Values given for the FBJ-MSV-induced tumour and FBJ-MSV- and SM-FeSV-transformed cells were multiplied by 0.46 and 0.18, respectively, to correct for the lesser extent of hybridization of the *v-onc* probes with the human *c-onc* homologues as compared with their murine (*fos*) or feline (*fms*) counterparts. The correction values were determined by comparing the levels of probe hybridization with human and mouse (*fos*) or human and cat (*fms*) denatured DNA. Since DNA-RNA hybrids are generally more stable than DNA-DNA hybrids, the corrected expression values indicated in the table represent approximations to the actual levels of expression in the virus-transformed cells. The deviations between estimated and actual values, however, are too low to influence the interpretation of the data. The 3T3-FMS2 cell line was established by transfection^{30,31} of SM-FeSV DNA²⁴ onto NIH/3T3 cells. A431 cells were obtained from Dr B. Sefton (Salk Institute) and Dr Gordon Gill (University of California, San Diego) and RS2 cells from Dr T. Curran (Salk Institute). BeWo, WISH, FL, AV-3, PA-1 and WI-38 cell lines were purchased from the American Type Culture Collection. All cell lines were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. ≤, Precise determination not possible as signal close to background; ND, not determined.

* Two or † four samples from four different individuals were analysed.

‡ Corrected values (see legend above).

extraembryonal tissues. The fact that similar patterns of expression have also been observed in mice⁵⁻⁸ indicates that the *c-fos* and *c-fms* proteins may have crucial roles in mammalian development. It will be intriguing to investigate whether these functions are correlated with the 'pseudo-malignant' growth properties²⁷ of the human trophoblast. The very high levels of *c-fos* expression in term chorion and amnion, however, would be more suggestive of an involvement of the *c-fos* protein in differentiation or transport processes in those embryo-derived cells whose primary functions are protection and nourishment of the human fetus. In contrast to *c-fos* and *c-fms*, the *c-Ki-ras*, *c-Ha-ras* and *c-myc* gene products appear to serve some more general function(s) in cellular metabolism, as their expression does not seem to be modulated in different tissues. The observation that these *c-onc* genes are transcriptionally active at significant levels in many, if not all human tissues, raises the possibility that structurally altered proteins from mutated *c-ras* or *c-myc* genes may have a role in the multi-stage process of neoplastic transformation of human cells. In this context it is noteworthy that *c-ras* and *c-myc* genes have been implicated in the induction of several human malignant tumours^{28,29}.

We thank T. Curran for reading of the manuscript and for RNA from a mouse osteosarcoma, M. Bishop, R. Ellis, D. Lowy, E. Scolnick and C. Sherr for recombinant plasmids containing *v-onc*-specific sequences, K. Benirschke for human tissue samples and D. Müller for technical assistance. R.M. was the recipient of a fellowship from the Deutsche Forschungs-

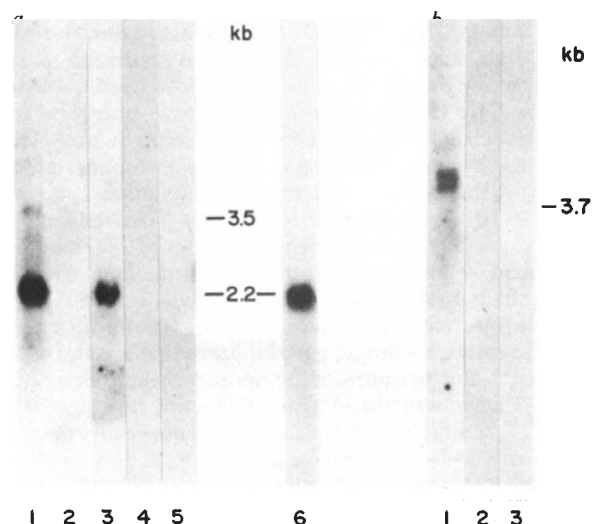


Fig. 1 Transcripts from *c-onc* genes in human tissues and cells. **a**, Transcripts from the *c-fos* gene. Lane 1, placenta; lane 2, lymph nodes; lane 3, BeWo cells; lane 4, WI-38 cells; lane 5, PA-1 cells; lane 6, chorion (for details see Table 1 legend). **b**, Transcripts from the *c-fms* gene. Lane 1, BeWo cells; lane 2, WI-38 cells; lane 3, lymph nodes. Twenty microgrammes of either total RNA (lane 6) or poly(A)⁺ RNA (lanes 1-5) were separated on 1.1% agarose gels, blotted onto nitrocellulose paper and hybridized to *v-onc*-specific probes as described previously^{5,6}.

gemeinschaft. This work was supported by grants from ACS and NIH.

Received 11 March; accepted 1 June 1983.

1. Bishop, J. M. & Varmus, H. E. in *RNA Tumor Viruses, The Molecular Biology of Tumor Viruses Pt III* (eds Weiss, A., Teich, N., Varmus, H. & Coffin, J.) 999–1108 (Cold Spring Harbor Laboratory, New York, 1982).
2. Chen, J. H. *J. Virol.* **36**, 162–170 (1980).
3. Shibuya, M., Hanafusa, H. & Balduzzi, P. L. *J. Virol.* **42**, 143–152 (1982).
4. Gonda, T. J., Sheiness, D. K. & Bishop, J. M. *Molec. cell. Biol.* **2**, 617–624 (1982).
5. Müller, R. *et al. Nature* **299**, 640–644 (1982).
6. Müller, R. *et al. Molec. cell. Biol.* **3**, 1062–1069 (1983).
7. Müller, R. *Cold Spring Harb. Cell Proliferation* **10** (in the press).
8. Müller, R., Verma, I. M. & Adamson, E. D. *EMBO J.* **2**, 679–684 (1983).
9. Westin, E. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2194–2198 (1982).
10. Westin, E. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2490–2494 (1982).
11. Van Beveren, C., van Straaten, F., Curran, T., Müller, R. & Verma, I. M. *Cell* **32**, 1241–1255 (1983).
12. Van Straaten, F., Müller, R., Curran, T., Van Beveren, C. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183–3187 (1983).
13. Curran, T., MacConnell, W. P., van Straaten, F. & Verma, I. M. *Molec. cell. Biol.* **3**, 914–921 (1983).
14. Hayflick, L. & Moorhead, P. S. *Expl Cell Res.* **25**, 585–621 (1961).
15. Zeuthen, J. *et al. Int. J. Cancer* **25**, 19–32 (1980).
16. Patillo, R. A. & Gey, G. O. *Cancer Res.* **28**, 1231–1236 (1968).
17. Hayflick, L. *Expl Cell Res.* **23**, 14–20 (1961).
18. Fogh, J. & Lund, R. *Proc. Soc. exp. Biol. Med.* **94**, 512–537 (1957).
19. Fabricant, R. N., DeLarco, J. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 565–569 (1977).
20. Rasheed, S. *et al. Cancer* **33**, 1027–1033 (1974).
21. Collins, S. J., Gallo, R. C. & Gallagher, R. E. *Nature* **270**, 347–349 (1977).
22. Adamson, E. D., Deller, M. J. & Warshaw, J. B. *Nature* **291**, 656–659 (1981).
23. Curran, T. & Teich, N. M. *Virology* **116**, 221–235 (1982).
24. Donner, L. *et al. J. Virol.* **41**, 489–500 (1982).
25. Chang, E. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 4848–4852 (1982).
26. Sheiness, D. & Bishop, J. M. *J. Virol.* **31**, 514–521 (1979).
27. Hertz, R. *Choriocarcinoma and Related Gestational Tumors in Women*, 1–12 (Raven, New York, 1978).
28. Cooper, G. M. *Science* **218**, 801–806 (1982).
29. Taub, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7837–7841 (1982).
30. Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456–467 (1973).
31. Corsaro, C. M. & Pearson, M. L. *Somat. Cell Genet.* **7**, 603–616 (1981).

Transforming activity of polyoma virus middle-T antigen probed by site-directed mutagenesis

Ben A. Oostra, Robert Harvey, Barry K. Ely, Alexander F. Markham* & Alan E. Smith

Biochemistry Division, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

*ICI Pharmaceuticals Division, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

The ability of polyoma virus to transform cells results primarily from the action of one of the virus-coded early proteins, called middle-T antigen^{1–3}. Middle-T has an associated tyrosine-specific protein kinase activity that can be measured *in vitro* and results in the phosphorylation of middle-T itself^{4–7}. Almost all mutants so far tested that lack the ability to transform cells, also lack associated kinase activity^{2,4}. Attempts to map within middle-T the tyrosine residue(s) that are phosphorylated *in vitro* suggest that a likely site of phosphorylation is tyrosine 315 (refs 8–10 and unpublished results). The amino acid sequence preceding Tyr 315 includes a tract of six contiguous glutamic acid residues and bears some homology with that preceding the tyrosine phosphorylated *in vivo* in pp60^{v-src}, the transforming protein of Rous sarcoma virus¹¹, and with a region in the polypeptide hormone, gastrin, preceding a tyrosine that is sulphated¹². Furthermore, although surprisingly large tracts of middle-T may be removed without affecting its transforming activity^{2,13–16}, mutants that lack the sequences corresponding to amino acids 311–318 inclusive are transformation defective. Because the likely site of phosphorylation, the homology with pp60^{v-src} and gastrin and the sequence apparently required for transformation all overlap, it has generally been accepted that this region of middle-T may form part of an essential region, possibly an active site on the protein¹⁷. Here we have used techniques of site-directed and site-specific mutagenesis to probe the sequence requirements in more detail. Contrary to expectation, the results obtained strongly suggest that Tyr 315 and conservation of the surrounding amino acid sequence are not essential for transformation.

A collection of mutants within the sequences removed from the deletion mutant *dl2208* (ref. 16) was constructed using deletion loop mutagenesis¹⁸. Plasmids containing the early region of polyoma virus DNA between the *Bam*HI and *Eco*RI restriction enzyme sites from wild-type (pAS101) and *dl2208* (pAS103) were digested with *Bam*HI and *Eco*RI, respectively, and heteroduplexes formed. Following treatment with sodium bisulphite and transfection into *Escherichia coli* K58, clones containing wild-type length plasmid DNA were selected and the appropriate region of the DNA was sequenced. The deletion in *dl2208* removes sequences corresponding to amino acids 298–316 (Fig. 1a). Because we were not interested in determining the effects of amino acids outside the glutamic acid-rich region, appropriate mutated plasmid DNAs were reconstructed using the *Pvu*II site (Fig. 1a), thus eliminating alterations in the sequences between amino acid residues 298 and 304.

One particularly important change that we wished to introduce was the conversion of tyrosine 315 to phenylalanine, such that phosphorylation of this residue would be prevented. This requires changing TAC to TTC in the viral DNA. Fortunately, this conversion results in the loss of an *Rsa*I restriction enzyme recognition site (GTAC); however, it cannot be achieved by bisulphite treatment.

A 19-residue oligodeoxynucleotide was synthesized using the solid phase phosphotriester method¹⁹. Its sequence was in the same sense as that of early messenger RNA and corresponded to nucleotides 1,170 and 1,188 of polyoma virus DNA, except that T rather than A was present at residue 1,178. Because we found some fragments of polyoma virus DNA to be unstable when inserted into bacteriophage M13, we did not use the primer-dependent oligonucleotide-directed mutagenesis procedure²⁰. Instead, as outlined in Fig. 1b, we formed a gapped, heteroduplex molecule from appropriate plasmid DNAs, to which the synthetic oligonucleotide could be hybridized. The desired mutant was readily detected following cleavage with *Rsa*I by the loss of fragments of 49 and 372 residues and the appearance of a new, longer fragment of 421 residues. The frequency of mutant production was at least 1 in 20.

All the plasmid DNAs harbouring appropriate mutations were reconstructed to give inserts capable of directing the synthesis of polyoma virus middle-T (as well as small-T and a fragment of large-T), and a fragment (*Ava*II/*Hpa*II) which spans the target region from selected, reconstructed plasmid DNAs was sequenced. In the 80 nucleotides from the *Pvu*II site, point mutations were detected only within the target area defined by the *Pvu*II site and the distal end of the deletion in *dl2208* DNA. No alterations outside the region sequenced were found on mapping the cleavage sites of restriction enzymes *Pvu*II, *Eco*RI, *Sac*I, *Hinf*I, *Ava*II and *Hpa*II (which flank or lie within the remainder of the 416 nucleotides from mutant plasmid DNAs that were cloned back into pAS101 DNA). The changes in nucleotide sequence of the mutants are shown in

Table 1 Tumour incidence in Fischer rats following injection of mutant plasmid transformed cells

	Expt 1	Expt 2		
Rat 1	0/3	—	0/3	—
pAS101	4/4	18	3/3	19
pAS131	4/4	13	4/4	21
pAS132	4/4	15	ND	—
pAS133	4/4	17	ND	—
pAS134	4/4	14	ND	—
pAS135	3/4	29	ND	—
pAS136	4/4	36	4/4	28
pAS137	4/4	26	ND	—

10⁶ Transformed cells were injected into the flank of 4-week-old Fischer rats, and the appearance of tumours was recorded. The table shows the number of rats with tumours and the average time (in days) before tumours were detected. No tumours were detected using Rat 1 cells after 60 days. ND, not determined. With pAS131, cells from independent transformed foci were used in the two experiments.

Fig. 1 Oligonucleotide-directed mutagenesis. *a*, The sequence of polyoma virus DNA between nucleotides 1,123 and 1,200 inclusive is shown together with the corresponding amino acid sequence of middle-T, and the locations of the restriction enzyme cleavage sites for *PvuII*, *RsaI* and *XhoII*. Also indicated are (part of) the sequences removed in deletion mutants *dl1020*, *dl2210*, *dl2208* and *dl23*. Open boxes indicate mutants that are transformation positive, hatched boxes are transformation defective. The amino acids apparently essential for transformation are boxed^{2,16}. *b*, Procedure for obtaining a gapped, heteroduplex molecule. Plasmid pAS101 DNA containing the 2,220-base pair (bp) *BamHI*-*EcoRI* fragment of wild-type polyoma virus DNA inserted in pAT153 was digested with *PvuII*, ligated with T4 DNA ligase and transfected in *E. coli* ED 8767 (*supE*, *supF*, *hsd* 5⁻, *met*⁻, *recA* 56). Clones were selected that contained only one *PvuII* site (pAS117). Plasmid DNA of pAS117 (5 µg) was (1) digested with *BamHI* and subsequently dephosphorylated with calf alkaline phosphatase, extracted with phenol and precipitated with ethanol; (2) digested with *PvuII* and *EcoRI* and the large 3,776-bp fragment isolated; (3) digested with *EcoRI* and *XhoII* (isolated according to J. Olsen, P. Meyers and R. Roberts, personal communication) and the 370-bp fragment of polyoma virus DNA isolated. The fragments from the three digestions were mixed together in 50 µl water. The DNA was denatured by adding 10 µl 1 M NaOH and the mixture was kept at room temperature for 10 min. Renaturation was accomplished by sequential addition of 500 µl water, 80 µl 0.5 M Tris-HCl pH 8 and 100 µl 0.1 M HCl followed by incubation at 65 °C for 2 h. The annealed products were ethanol precipitated and resuspended in 10 µl 1 mM Tris-HCl pH 8, 0.1 mM EDTA. Half of the annealed DNA was combined with 100 pmol of 5' phosphorylated oligonucleotide GGAGGAGTTCATGCCAATG (containing a mismatch at position 9, where an A is replaced by a T) in 10 µl 50 mM NaCl, 6.6 mM Tris-HCl pH 7.6, 6.6 mM MgCl₂, 1 mM dithiothreitol (DTT). To this was added 5 µl of a solution which contained the four deoxyribonucleotides at 2.5 mM, plus 2.5 mM ATP, 50 mM NaCl, 5.5 mM Tris-HCl pH 7.6, 6.6 mM MgCl₂, 1 mM DTT. After addition of 1 U large fragment (Klenow) DNA polymerase I and 20 U T4 DNA ligase, the mixture was incubated for 15 min at 0 °C and for 60 min at 18 °C. Portions of this reaction mixture were used to transfect *E. coli* ED 8767. DNA from the resulting colonies was tested for the loss of an *RsaI* site (GTAC). From one of the clones with the desired mutation (pAS113), DNA was digested with *PvuII* and *EcoRI* and the 416-bp fragment was isolated. This fragment was used to replace the 416-bp fragment in pAS101 to form pAS131. B, *BamHI*; P, *PvuII*; X, *XhoII*; E, *EcoRI* cleavage sites.

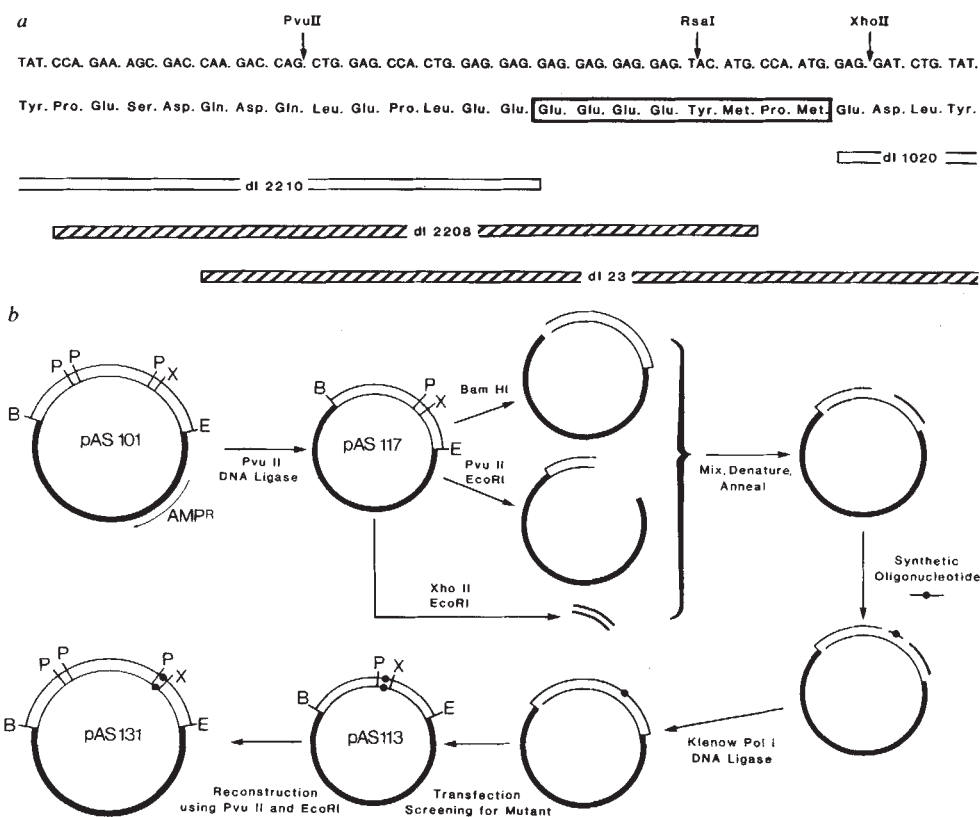


Fig. 2a and the corresponding amino acids are indicated in **Fig. 2b**. Some of the mutants include single amino acid changes, whereas others have two, three or more changes. If the hypothesis being tested is correct we would expect some of the mutants to be defective in transformation.

Purified plasmid DNAs were introduced into Rat 1 cells using calcium phosphate transfection and middle-T transformed cells were selected as dense foci of cells overgrowing the cell monolayer²¹. In these conditions middle-T is sufficient to transform Rat 1 cells³. Wild-type plasmid DNA (pAS101) gave ~90 foci after transfection of 5 µg plasmid DNA onto ~10⁶ cells. Foci of transformed cells were detected after ~14 days. Parallel experiments using plasmid DNA from the parent deletion mutant (pAS103) gave no foci after 30 days. When mutant plasmid DNAs were tested in the same way, all gave approximately the same number of foci per µg added DNA. The time of appearance of foci was sometimes later by 1–3 days with some of the mutants (pAS131 and 136) but foci of transformed cells could be picked readily in all cases by 19 days after transfection. No differences in morphology between wild-type and mutant transformed cells were detected.

Transformed cell lines were grown up from a number of randomly selected foci and tested for anchorage-independent growth in agar²². Different numbers of transformed cells were suspended in semi-solid medium (0.3% agar) and the appearance of colonies of transformed cells was noted. In these conditions Rat 1 cells did not proliferate, whereas more than 90% of the suspended cells gave rise to colonies when pAS101

<i>a</i>	
PAS101	GAG.GAG.GAG.GAG.GAG.GAG.TAC.ATG.CCA.ATG
PAS131	—————T—————
PAS132	—————A—A—A—————
PAS133	—————A—————
PAS134	—————A—A—A—————
PAS135	—————A—————
PAS136	A—————A—A—A—A—————
PAS137	—————A—A—————A—A—————
<i>b</i>	
PAS101	GLU.GLU.GLU.GLU.GLU.GLU.TYR.MET.PRO.MET
PAS131	—————PHE—————
PAS132	—————ILE—————
PAS133	—————LYS—ILE—————
PAS134	—————LYS.LYS—————
PAS135	—————LYS—————
PAS136	LYS—————LYS.LYS.LYS—————
PAS137	—————LYS—————ILE—————

Fig. 2 A collection of mutants around Tyr 315. *a*, Changes in the DNA sequence of the various mutants; *b*, the corresponding changes in the amino acid sequence of middle-T.

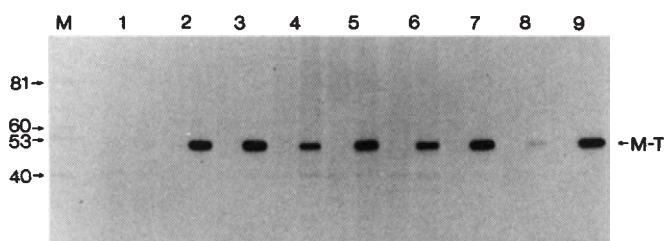


Fig. 3 Middle-T associated kinase activity in mutant transformed cells. Extracts from $\sim 5 \times 10^5$ transformed cells were immunoprecipitated using hamster control and anti-T serum, washed and assayed for kinase activity in conditions described in detail previously^{4,7,27}. The labelled products were separated on a 10% polyacrylamide gel, which was dried and autoradiographed for 16 h. Extracts used were from: 1, Rat 1 cells, and cells transformed by: 2, pAS101; 3, pAS131; 4, pAS132; 5, pAS133; 6, pAS134; 7, pAS135; 8, pAS136; 9, pAS137. Alternate lanes show the reaction with hamster control and anti-T serum. In the gel conditions used middle-T (M-T) does not separate into the two bands reported by others^{9,25}. M, molecular weight ($\times 10^{-3}$).

plasmid transformed cells were used. No significant difference from wild-type behaviour was detected with the mutant transformed cells, either in the number or size of colonies detected or their time of appearance.

The mutant plasmid transformed cells were also injected into young adult Fischer rats and the appearance of tumours was monitored. Tumours were first detected using wild-type plasmid pAS101 DNA transformed cells about 18 days after injection. As shown in Table 1, tumours were obtained with all the mutant plasmid transformed cells and the majority (pAS131, 131, 133, 134) were first detected between 13 and 17 days after injection. In some cases, the latent period was slightly longer, and in one case (pAS136) 36 days. In some cases, cell lines were derived from the tumours and used in parallel with the plasmid transformed cell lines.

Extracts were made from the different mutant transformed cell lines immunoprecipitated with anti-T serum, and tested for kinase activity^{4,7}. Figure 3 shows that middle-T was labelled in the reaction using hamster anti-T serum and extracts of wild-type plasmid transformed cells but not using control hamster serum or extracts of Rat 1 cells. Middle-T was also labelled in extracts of all the transformed cells, including those transformed by the mutants lacking Tyr 315 (pAS131). The amount of ^{32}P -labelled middle-T was significantly less in extracts of pAS136 transformed cells. In parallel experiments (data not shown), the transformed cells were labelled with ^{35}S -methionine and the labelled proteins immunoprecipitated with anti-T serum. All the transformed cells contained approximately equivalent amounts of middle-T except pAS136, which contained about 90% less.

Extracts of cells derived from tumours caused by mutant transformed cells (pAS131–136) were also used for kinase assays. In all cases the results obtained were indistinguishable from those obtained with extracts of transformed cells. The experiments reported in Fig. 3 were also repeated using a particular rat anti-T serum which when used in the kinase reaction leads to the phosphorylation of the rat immunoglobulin but not middle-T^{4,7} and a rabbit antiserum raised against a synthetic peptide corresponding to the carboxy-terminal six amino acids of middle-T²³. The results obtained using these sera paralleled those obtained using hamster serum. Phosphoamino acid analysis of the ^{32}P -labelled mutant middle-T species showed that in all cases the acceptor was still tyrosine (data not shown).

These results show that, as expected, all the transformed cell lines contain middle-T. They also indicate that in each case the mutant middle-T retains the ability to accept phosphate and has associated kinase activity, as measured by the ability to phosphorylate an exogenous acceptor (rat immunoglobulin). The most striking result, however, is the finding that extracts of pAS131 transformed cells still retain the ability to accept wild-type levels of phosphate in spite of having lost the tyrosine

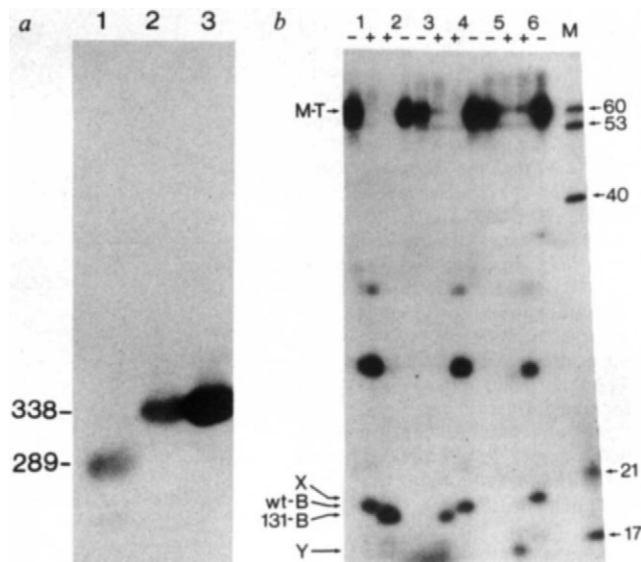


Fig. 4 Analysis of DNA and protein from mutant transformed cells. **a**, High molecular weight DNA from transformed cells was isolated and digested with *Hae*III. After separation on an agarose gel, *Hae*III fragments in the range 600–800 nucleotides were isolated, digested with *Rsa*I, separated on an agarose gel and transferred to nitrocellulose paper. The filters were hybridized with the 289-bp *Rsa*I fragment of pAS101 which was labelled by nick-translation. After washing and drying, the filter was autoradiographed. DNA was isolated from cells transformed with: lane 1, pAS101; lane 2, pAS131; lane 3, pAS131 (a different cell line from that shown in lane 2). **b**, Extracts from $\sim 2 \times 10^6$ cells were immunoprecipitated, washed, incubated with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and separated on preparative scale 10% polyacrylamide gels. The gels were wrapped in plastic film and autoradiographed without drying for 16 h. ^{32}P -middle-T was excised, homogenized to an even paste and incubated with 0.02 mg ml^{-1} *S. aureus* V8 protease for 1 h at 37°C as described previously⁷. Samples were electrophoresed on a second, 20-cm long, 15% polyacrylamide gel. Middle-T was from: 2, a cell line from a tumour induced by pAS131 transformed cells; cells transformed by: 1, pAS101; 3, pAS131; 4, pAS101; 5, pAS134; 6, pAS133. + And – indicate samples incubated with or without enzyme. M indicates markers with molecular weights as shown ($\times 10^{-3}$). Wt-B indicates fragment B from pAS101 wild-type transformed cells, 131B the equivalent fragment from pAS131 transformed or tumour cells, and X and Y the altered fragment from pAS133 and pAS134, respectively.

reported to be the major acceptor site in wild-type middle-T^{9,10}. The most obvious explanation of this result is that one or more other tyrosine residues are being phosphorylated in the mutant middle-T species.

Because the phenotype of the mutant transformed cells was virtually wild type, we wished particularly to confirm that the pAS131 transformed cells were indeed transformed by mutant DNA rather than a revertant molecule. To do this, high molecular weight DNA was isolated and screened for the presence of the *Rsa*I site which includes the codon for Tyr 315. Figure 4a shows that, as expected, wild-type transformed cells contain a 289-nucleotide *Rsa*I fragment from this region. By contrast, two independent pAS131 transformed cell lines contain a 338-nucleotide fragment. This confirms that the cells do not contain wild-type polyoma DNA and strongly suggests that they are transformed by pAS131 DNA.

Partial proteolysis of ^{32}P -middle-T labelled in the kinase reaction gives a reproducible pattern of fragments following digestion with *Staphylococcus aureus* V8 protease⁷. The digestion products include a prominent labelled fragment (called B) of $\sim 18,000$ molecular weight, which is thought to include the carboxy-terminal region of middle-T and may result from cleavage of the protein in the region of the glutamic acid tract (refs 7,9 and unpublished results). Figure 4b shows partial proteolysis fingerprints of middle-T from extracts of pAS101 wild-

type transformed cells and from pAS131 mutant transformed cells. Fragment B is labelled in pAS131 middle-T, but it has a slightly faster mobility than the wild-type fragment. Because the pattern is not wild type, this experiment confirms that the protein expressed in pAS131 transformed cells is indeed a mutant middle-T species. Experiments to establish the reason for the altered mobility of the fragment are in progress. Analysis of the other mutant middle-T species by partial proteolysis mapping shows that slightly altered patterns are obtained with pAS133 and 134 (Fig. 4), whereas in other cases (pAS137), the cleavage pattern is indistinguishable from wild type.

The experiments described here tested the hypothesis that the sequence Glu-Glu-Glu-Glu-Tyr-Met is required for transformation and, in particular, that Tyr 315 is essential. The results indicate that this hypothesis is incorrect. The mutant plasmid DNAs all transformed cells and when transformed foci were grown up, all the cells produced colonies in agar and tumours in rats. In the only case where the latent period was significantly longer (pAS136), the cells contained only about 10% of the middle-T detected in the other lines.

It is possible that the sequence around Tyr 315 is important for reasons not related to phosphorylation but rather, for example, because it is a recognition site for another protein. Our data also seem to rule out this hypothesis. Mutant pAS136 has a sequence Lys-Glu-Lys-Lys-Lys-Glu-Tyr rather than Glu-Glu-Glu-Glu-Glu-Glu-Tyr. In spite of this radical change in sequence, the mutant still transforms and causes tumours. Very recently, another deletion mutant (called *dl17*) capable of transforming has been reported in this region²⁴. The mutant protein produced lacks the glutamic acid tract preceding Tyr 315, but has instead the sequence Glu-Ser-Asp-Gln-Glu-Tyr, which is even more closely homologous to the pp60^{v-src} sequence (Glu-Asp-Asn-Glu-Tyr) than is wild-type middle-T. Interpretation of results obtained with this mutant is therefore difficult. However, the data reported here indicate that the homology with pp60^{v-src} is unimportant. Taken together, all the data suggest that sequence conservation of the glutamic acid-rich region is not essential for transformation. The present findings do not explain why previous work with the deletion mutants indicated that this region was important.

Our data also suggest that the tyrosine kinase can phosphorylate more than one site in middle-T. Thus, although the protein may be phosphorylated primarily on Tyr 315 in normal conditions, in pAS131 middle-T this cannot be the case. This implies that the kinase is able to use more than one site. Such a conclusion has already been suggested by the findings that *dl23* middle-T, which lacks Tyr 315 and 322 and *dl2208* middle-T, which lacks Tyr 315, are both weakly phosphorylated on tyrosine in the *in vitro* kinase reaction (refs 7, 25 and G. Belsham, personal communication). Further experiments to map the additional phosphorylation sites are in progress.

Although the results reported here show that phosphorylation of Tyr 315 is not essential for transformation, we cannot exclude a model that predicts that phosphorylation of tyrosine is required, but that the location of the tyrosine is irrelevant. We consider this model unlikely. It is perhaps significant that most of the other tyrosine residues in the unique carboxy-terminal portion of middle-T can be removed in different deletion mutants without affecting transformation (for example, 258 in *dl8* (ref. 13), 289 and 297 in *dl45* (ref. 14) and 322 in *dl1013* (ref. 15)). Furthermore, middle-T isolated from cells contains very little phosphate, and although there is one report that this includes phosphotyrosine²⁶, no attempts have been made to map the site.

Any model for transformation involving phosphorylation of middle-T concentrates on the action of the tyrosine kinase on middle-T. Although such an activity can be measured *in vitro*, it may be unimportant in cells. An alternative explanation is that the phosphorylation of middle-T reflects its presence in a complex with a cellular tyrosine kinase. If, as previously suggested²⁷, the associated kinase is pp60^{c-src}, the cellular homologue of the transforming protein of Rous sarcoma virus,

it is quite conceivable that the normal reactions catalysed by pp60^{c-src} are altered by its interaction with middle-T and that this results in transformation. Our future experiments will test this model.

After submission of this manuscript, Snyder *et al.*²⁸ reported that Tyr 416 of pp60^{v-src}, which is normally phosphorylated, can be replaced by phenylalanine without any measurable effects on either transforming ability or tyrosine protein kinase activity.

We thank Peter Gillett for assistance with some of these experiments, Mike Edge for help in the purification of the synthetic oligonucleotide, Graham Belsham, Dan Kalderon, Sara Courtneidge and Bill Richardson for helpful comments throughout the course of this work, and Lydia Pearson for help with typing. B.A.O. was supported by the Bruno Mendel Fellowship of the Royal Society and by EMBO.

Received 18 March; accepted 16 June 1983.

1. Tooze, J. *DNA Tumor Viruses* 2nd edn, Pt 2 (Cold Spring Harbor Laboratory, New York, 1980).
2. Smith, A. E. & Ely, B. K. in *Advances in Viral Oncology* Vol. 3 (ed. Klein, G.) 3-30 (Raven, New York, 1983).
3. Treisman, R., Novak, U., Favaloro, J. & Kamen, R. *Nature* **292**, 595-600 (1981).
4. Smith, A. E., Smith, R., Griffin, B. & Fried, M. *Cell* **18**, 915-924 (1979).
5. Eckhart, W., Hutchinson, M. A. & Hunter, T. *Cell* **18**, 925-933 (1979).
6. Schaffhausen, B. S. & Benjamin, T. L. *Cell* **18**, 935-946 (1979).
7. Smith, A. E., Fried, M., Ito, Y., Spurr, N. & Smith, R. *Cold Spring Harb. Symp. quant. Biol.* **44**, 141-147 (1980).
8. Smith, A. E. in *Protein Phosphorylation and Bioregulation* (eds Thomas, G., Podesta, E. & Gordon, J.) 219-228 (Karger, Basel, 1980).
9. Schaffhausen, B. & Benjamin, T. *J. Virol.* **40**, 184-196 (1981).
10. Schaffhausen, B. & Benjamin, T. in *Cold Spring Harbor Conf. Cell Proliferation* Vol. 8 (eds Rosen, O. M. & Krebs, E. G.) 1281-1298 (1981).
11. Smart, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6013-6017 (1981).
12. Baldwin, G. S. *FEBS Lett.* **137**, 1-5 (1982).
13. Griffin, B. E. & Maddock, C. J. *J. Virol.* **31**, 645-656 (1979).
14. Bendig, M., Thomas, T. & Folk, W. *J. Virol.* **33**, 1215-1219 (1980).
15. Magnusson, G., Nilsson, M. J., Dilworth, S. M. & Smolar, N. *J. Virol.* **39**, 673-683 (1981).
16. Nilsson, S., Tyndall, C. & Magnusson, G. *J. Virol.* **46**, 284-287 (1983).
17. Patschinsky, T., Hunter, T., Esch, F. S., Cooper, J. A. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 973-977 (1982).
18. Kalderon, D., Oostra, B. A., Ely, B. K. & Smith, A. E. *Nucleic Acids Res.* **10**, 5161-5171 (1982).
19. Edge, M. D. *et al. Nature* **292**, 756-761 (1981).
20. Shortle, D., DiMaio, D. & Nathans, D. *A. Rev. Genet.* **15**, 265-294 (1981).
21. Graham, F. L. & van der Eb, A. *Virology* **52**, 456-458 (1973).
22. MacPherson, I. & Montagnier, L. *Virology* **23**, 291-294 (1964).
23. Walter, G., Hutchinson, M. A., Hunter, T. & Eckhart, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4882-4886 (1981).
24. Ding, D., Dilworth, S. M. & Griffin, B. E. *J. Virol.* **44**, 1080-1083 (1982).
25. Dilworth, S. *EMBO J.* **1**, 1319-1328 (1982).
26. Segawa, K. & Ito, Y. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6812-6816 (1982).
27. Courtneidge, S. A. & Smith, A. E. *Nature* **303**, 435-439 (1983).
28. Snyder, M. A., Bishop, J. M., Colby, W. W. & Levinson, D. A. *Cell* **32**, 891-901 (1983).

Critical role of an eight-amino acid sequence of VP1 in neutralization of poliovirus type 3

D. M. A. Evans*, P. D. Minor*, G. S. Schild* & J. W. Almond†

* National Institute for Biological Standards and Control, Holly Hill, London NW3 6RB, UK

† Department of Microbiology, University of Leicester, Leicester LE1 7RH, UK

The three serotypes of poliovirus are members of the picornaviridae, a group of viruses which cause a variety of diseases of considerable importance in man and animals. We have previously used antigenic mutants resistant to neutralizing monoclonal antibodies to identify a single antigenic site for the neutralization of poliovirus type 3 (ref. 1). Evidence based on oligonucleotide mapping suggested that this site corresponded largely to one physicochemical region of the capsid protein viral polypeptide 1 (VP1). We now present conclusive evidence that most of the mutations conferring resistance to neutralization are confined to an eight-amino acid region of VP1, specified by a sequence of viral RNA 277-300 bases from the start of the region coding for VP1. These data strongly suggest that this small region constitutes a major antigenic site involved in virus neutralization and they provide the most detailed information currently available on the antigenic site of a human virus.

The preparation of neutralizing monoclonal antibodies specific for type 3 poliovirus used in these studies is reported elsewhere²⁻⁴. As an extension of our previous studies, mutants of a single parental poliovirus type 3 strain (P3-Leon-USA-1937) were selected in the presence of several individual virus neutralizing monoclonal antibodies; 16 mutant groups, including the 10 we have previously described¹, were identified by their distinct patterns of resistance against 11 monoclonal antibodies. Their reaction patterns are presented in Table 1 which also identifies the monoclonal antibodies used to select the various groups of mutant viruses. The data presented here add further weight to our earlier conclusion that all the mutations are localized in the same antigenic site. For the purpose of this analysis, an antigenic site is defined as an area of the virus where a mutation affecting the binding of one monoclonal antibody affects the reactions of others directed against the same site¹. Such an operationally-defined site need not necessarily be made up of a contiguous sequence of amino acids⁵. Our previous data implicated modifications in a limited region of VP1 in the generation of new antigenic configurations. The RNA of representative strains from each mutant group was sequenced to define more precisely the positions of amino acid substitutions in VP1 which conferred resistance. This was done by the dideoxy sequencing method⁵ using a primer restriction fragment prepared from cloned poliovirus cDNA⁶.

This sequence method produced unambiguous results. Consequently, the RNA used must have been largely homogeneous; a mixed population of RNAs would have resulted in unreadable sequence ladders. A sequence of at least 60 bases, and in most cases up to 150 bases, was identified for each mutant. The sequences of all the mutants were the same except for the single point mutations shown in Fig. 1. Where mutations were detected there was only one per mutant and all would result in amino acid substitutions.

Base changes were detected in viruses of 15 of the 16 mutant groups and were concentrated into a single region of only eight codons (Fig. 1). Only one mutant group (group 9) had an amino acid substitution at position 4 (glutamic acid to glycine). The amino acids at positions 5 and 6 were conserved in all the mutant groups. The remaining mutations were distributed between positions 7 and 11. For example, mutants of groups 1, 2 and 14 showed different point mutations in position 7 and at least two distinct mutations were located in each of positions 8 to 11. Furthermore, different mutations which occurred in the same position resulted in different patterns of resistance to neutralization. Single base changes at the six variable codons could generate 33 possible different mutants. We have therefore isolated nearly half of the total number of theoretically possible mutants.

We have attempted to identify the precise amino acids required for the neutralization of poliovirus by the specific monoclonal antibodies studied here. An example of this approach is given in Table 2 for the antibody 25-1-14. Mutants were classified according to whether they were neutralized or not by this antibody. It can be seen that all substitutions at positions 4, 7 and 9 produced viruses which this antibody did not neutralize. Conversely, substitutions at position 10 had no effect on neutralization. In addition, substitutions in position 8, by amino acids lacking hydroxyl groups, were associated with lack of neutralization. The effect of substitutions at position 11 was ambiguous. We therefore conclude that antibody 25-1-14 required glutamic acid at position 4, threonine at position 7, threonine or serine at position 8 and arginine at position 9 for its neutralizing activity. Table 3 summarizes the results of a similar analysis of all 11 monoclonal antibodies used for the analysis of the mutants. It can be seen that the amino acids at positions 4 and 9 in particular effected neutralization of the mutant viruses. In addition, six antibodies neutralized mutants having a serine in place of the threonine at position 8 while not neutralizing viruses having other substitutions at this position. A seventh antibody (208) did not neutralize viruses with any substitution at this position. This is consistent with the view

that hydroxy amino acids are important in binding of neutralizing antibody to the virus. The majority of the required amino acids were polar in nature as expected for the residues participating in the reaction of antibody with antigen. This supports the view that the eight-amino acid sequence on VP1 represents the site on the virus to which neutralizing antibodies bind.

It was previously reported¹ that all members of group 5 had a mutation in the nonstructural part of the genome coding for the polymerase. The data shown here indicate that members of this group were double mutants as they also possess a mutation in position 11 (see Fig. 1). In only one of the groups, group 10, was no mutation detected in the 150 bases sequenced by primer extension, indicating that in this case the mutation was elsewhere in the genome. Nevertheless, this mutation might be brought into the same single antigenic site in the tertiary or quaternary structure of the capsid. Such a feature is not without precedent. The haemagglutinin molecule of influenza A virus has been found to contain five operationally-defined antigenic sites all of which are coded for by non-contiguous portions of the primary gene sequence⁵. Other reports suggest the existence of multiple antigenic sites for poliovirus type 1 (refs 7, 8).

The amino acid sequence for poliovirus type 1 (Mahoney strain) in the region we propose for the antigenic site of type 3 virus (positions 4-11) is also shown in Fig. 1. There are only two positions within this region which are conserved between types 3 and 1 and one of these is not conserved between the wild-type 1 vaccine progenitor strain (Mahoney) and the Sabin type 1 vaccine strain, LSC. In contrast, the two amino acids at

	1	2	3	4	5	6	7	8	9	10	11	12
Type 1 (Mahoney)	GUG Val	GAU Asp	AAC Asn	CCA Pro	CCU Ala	UCG Ser	ACC Thr	ACG Thr	AAU Asn	AAG Lys	GAU Asp	AAG Lys
Type 3 (Leon)	GUG Val	GAC Asp	AAU Asn	GAA Glu	CAA Gln	CCA Pro	ACC Thr	ACC Thr	CGG Arg	CCA Ala	CAG Gln	AAA Lys
Mutant group												
1							AUC Ile					
2							GCC Ala					
3								AAC Asn				
4									CAG Gln			
5											CUG Leu	
6											CCG Pro	
7										ACA Thr		
8								UCC Ser				
9				GGA Gly								
10												
11										GUA Val		
12											CGG Arg	
13											CAC His	
14							AAC Asn					
15									UGG Trp			
16							AUC Ile					

Fig. 1 The RNA of representative strains of each mutant group derived from Leon type 3 poliovirus was sequenced in the region 393-240 bases downstream from the 5' end of the VP1 coding region of the genome. The dideoxy sequencing method was used, using a restriction fragment primer prepared from cloned poliovirus cDNA. The primer was prepared by digesting plasmid DNA with *Eco*RI and *Sph*I and labelled by incubation with 1 unit of Klenow fragment from DNA polymerase 1 and 100 μ Ci of [α -³²P]dATP (3,000 Ci mmol⁻¹; Amersham). The products were denatured and loaded onto a preparative slab polyacrylamide gel. After electrophoresis the 47-base restriction fragment was detected by autoradiography, excised and eluted from the gel. The primer was purified by exclusion chromatography and then annealed onto the virion RNA and incubated with dideoxyribonucleotides and deoxyribonucleotides in appropriate proportions together with 5 units of avian myeloblast reverse transcriptase (Life Sciences Inc.). Following incubation at 37°C for 25 min the reaction products were denatured and loaded onto sequencing gels⁵. After electrophoresis the gels were fixed in 10% acetic acid, washed, dried and subjected to autoradiography.

Table 1 Reaction of neutralizing monoclonal antibodies with antigenic mutants

Mutant	25-1-14	25-4-12	27-4-4	199	194	134	208	175	204	197	165	Selecting monoclonal antibodies
1	r	r		r						r		25-1-24, 24-4-12, = 5-5-5, 27-4-4
2	r	r		r	r					r	r	25-1-14, 25-4-12, 199, 165, 25-5-5
3	r	r	r	r	r		r	r	r	r	r	25-4-12, 27-4-4, 175, 204, 25-5-5
4	r	r	r	r	r		r	r		r	r	25-4-12, 27-4-4, 204, 165, 132
5				r	r	r	r	r			r	199, 134, 132, 165, 175, 204
6				r	r	r	r	r			r	199, 165, 132
7				r	r	r	r	r	r		r	134, 165
8			r			r	r			r		27-4-4
9	r	r	r	r	r	r	r	r		r	r	27-4-4, 199, 175, 204
10	r		r	r	r	r	r	r	r	r	r	27-4-4, 134, 175
11				r	r		r	r		r	r	199, 175, 204, 165
12		r		r	r	r	r	r			r	175
13	r	r		r		r	r	r			r	175, 204
14	r	r						r		r	r	175
15	r	r	r	r	r			r	r	r	r	175, 204
16	r	r		r	r		r		r		r	199, 204, 165

Reaction was assessed by neutralization of 5×10^4 TCID₅₀ (50% tissue culture infective dose) units of type 3 poliovirus by a 1:10 dilution of antibody ascitic fluid. The wild type was neutralized by all antibodies. r, Resistant.

Table 2 The amino acid recognition requirement of monoclonal antibody 25-1-14 for neutralization of mutant viruses

Amino acids at position:													
Parental sequence	1	2	3	4	5	6	7	8	9	10	11	12	
	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	Ala	Gln	Lys	
a	Gp5	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	Ala	<u>Leu</u>	Lys
	Gp6	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	Ala	<u>Pro</u>	Lys
	Gp7	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	<u>Thr</u>	<u>Gln</u>	Lys
	Gp8	Val	Asp	Asn	Glu	Gln	Pro	Thr	<u>Ser</u>	Arg	Ala	Gln	Lys
	Gp11	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	<u>Val</u>	Gln	Lys
	Gp12	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	<u>Ala</u>	<u>Arg</u>	Lys
b	Gp1	Val	Asp	Asn	Glu	Gln	Pro	<u>Ile</u>	Thr	Arg	Ala	Gln	Lys
	Gp2	Val	Asp	Asn	Glu	Gln	Pro	<u>Ala</u>	Thr	Arg	Ala	Gln	Lys
	Gp3	Val	Asp	Asn	Glu	Gln	Pro	<u>Thr</u>	<u>Asn</u>	Arg	Ala	Gln	Lys
	Gp4	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	<u>Gln</u>	Ala	Gln	Lys
	Gp9	Val	Asp	Asn	<u>Gly</u>	Gln	Pro	Thr	Thr	Arg	Ala	Gln	Lys
	Gp13	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	Ala	<u>His</u>	Lyn
	Gp14	Val	Asp	Asn	Glu	Gln	Pro	<u>Asn</u>	Thr	Arg	Ala	<u>Gln</u>	Lys
	Gp15	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	<u>Trp</u>	Ala	Gln	Lys
	Gp16	Val	Asp	Asn	Glu	Gln	Pro	Thr	<u>Ile</u>	Arg	Ala	Gln	Lys

a, Amino acid substitutions in the proposed antigenic site of type 3 poliovirus mutants derived from Leon which are neutralized by monoclonal antibody 25-1-14. b, Amino acid substitutions in the antigenic site of mutant viruses which result in them not being neutralized by 25-1-14.

Table 3 Position of amino acids necessary for effective neutralization

Monoclonal antibody	Position no. and parental amino acid sequence:											
	1	2	3	4	5	6	7	8	9	10	11	12
	Val	Asp	Asn	Glu	Gln	Pro	Thr	Thr	Arg	Ala	Gln	Lys
25-1-14				+			+	+	+	-	?	
25-4-12				+			+	+	+	-	?	
27-4-4				+			-	?	+	-	-	
199				+			?	+	+	+	+	
194				+			?	+	+	+	?	
134				+			-	?	-	?	+	
208				+			-	+	?	+	+	
175				+			?	?	?	+	?	
204				-			-	+	+	?	+	
197				+			+	?	+	?	?	
165				+			?	+	+	+	+	
No. of monoclonal antibodies requiring parental amino acid at the stated position				10			3		8	5	4	

Blank represents no mutations detected at this site, that is, these amino acids were conserved throughout. +, The monoclonal antibodies have an unambiguous requirement for parental amino acid at this position. -, No positive requirement for the parental amino acid at this position. ?, Substitution of an amino acid at this position had an ambiguous effect on virus neutralization.

* Antibodies fail to react with viruses with any substitution other than a serine at this position.

positions 5 and 6 (glutamine and proline) were conserved in all mutants derived from the Leon type 3 virus, the Sabin type 3 vaccine and a virulent type 3 vaccine-associated strain (119) (A. J. Cann, unpublished data), however, they are not conserved between polioviruses of types 3 and 1. The same two positions are conserved between the wild-type 1 vaccine progenitor strain (Mahoney) and the Sabin type 1 vaccine strain (LSC). Thus, the proposed antigenic site shows low overall sequence homology between these two types of poliovirus. The significance of these observations with respect to type specificity is not clear. Nevertheless, it is possible that for poliovirus type 3, the conserved amino acids in positions 5 and 6 contribute to the type specificity of the neutralization reaction. However, confirmation of this will require the sequencing of the antigenic site of a range of wild-type strains of independent origin.

We thank the McAlpine Trust for its support.

Received 19 April; accepted 1 June 1983.

1. Minor, P. D. *et al.* *Nature* **301**, 674–679 (1983).
2. Ferguson, M., Minor, P. D., Magrath, D. I., Spitz, M. & Schild, G. C. *Rep. Int. Symp. on Poliomyelitis*, Washington, 1983 (in the press).
3. Ferguson, M. *et al.* *Lancet* **ii**, 122–124 (1982).
4. Minor, P. D. *J. gen. Virol.* **59**, 307–317 (1982).
5. Caton, A., Brownlee, G. G., Yewdell, J. W. & Gerhard, W. *Cell* **31**, 417–427 (1982).
6. Cann, A. J. *et al.* *Nucleic Acids Res.* (in the press).
7. Emini, E. A., Jameson, B. A., Lewis, A. J., Larsen, G. R. & Wimmer, E. *J. Virol.* **43**, 997–1005 (1982).
8. Emini, E. A., Shaw-Yi, Kao, Lewis, A. J. & Wimmer, E. *J. Virol.* (in the press).

Ca²⁺ channel modulation by 8-bromocyclic AMP in cultured heart cells

A. B. Cachelin, J. E. de Peyer, S. Kokubun & H. Reuter*

Department of Pharmacology, University of Berne, 3010 Berne, Switzerland

Modulation of ion channels is of increasing interest as it is an important step in the regulation of cellular functions^{1,2}. We have analysed the effect of 8-bromocyclic AMP on Ca²⁺ channels in cultured cardiac cells by the patch-clamp method³ and report here that there was a large increase in the probability of opening of the channels. On the basis of a recently proposed kinetic reaction scheme⁴ we suggest that cyclic AMP-dependent phosphorylation of Ca²⁺ channels primarily promotes the forward rate constants which lead to the open state of a Ca²⁺ channel during depolarization.

Primary cardiac cell cultures were prepared from hearts of neonatal rats by a method described elsewhere⁵. Experiments were performed on differentiated single myocytes 2–3 days after the start of the tissue culture. Cells grown on coverslips were placed in a chamber superfused with saline of the following composition (mM): NaCl 137; KCl 5.4; MgCl₂ 1.0; CaCl₂ 0.02; HEPES buffer 10; glucose 10, pH 7.4; temperature 24–26°C. 8-Bromocyclic AMP was added to the solution to yield final concentrations of 5 or 10 μ M. The recording pipette contained 96 mM BaCl₂ solution with 20 μ M tetrodotoxin, buffered to pH 7.4 with 10 mM HEPES. Single-channel currents, with Ba²⁺ as charge carriers, were recorded from cell-attached membrane patches having seal resistances between 10 and 100 G Ω ^{3,6}. Digitized records (sampling interval 0.2 ms) from six experiments were analysed in detail by means of a minicomputer (PDP 11/04). Although the cells often fired spontaneous action potentials, corresponding contractions were small due to the low extracellular Ca²⁺ concentration.

Within ~10 min after addition to the bathing solution, 8-bromocyclic AMP (10 μ M) caused reproducible and reversible changes in single-channel current traces (five experiments). Smaller concentrations of the nucleotide (5 μ M) had no clear effects (two experiments). Usually, we started with a control period of ~15 min, measured the effect of nucleotide addition

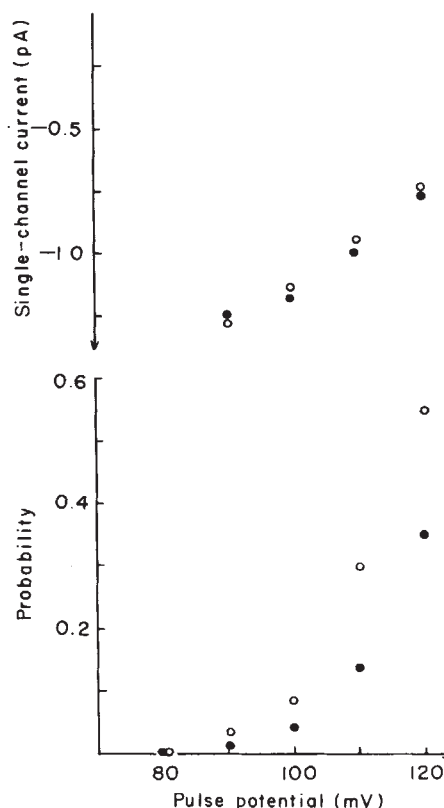


Fig. 1 Voltage dependence of single Ca²⁺ channel current (upper part) in the absence (closed circles) and presence (open circles) of 8-bromocyclic AMP. Open-channel current, i , was estimated by eye (see Fig. 3A) and each circle represents the mean of several hundred openings. Voltage dependence of opening probability (lower part) was calculated from $P = I/Ni$, where I is the time-averaged current flowing through open Ca²⁺ channels for 100 ms, and N is the number of functioning Ca²⁺ channels ($N = 1$). Pulse potentials refer to depolarizing clamp steps from a holding potential, V_H , 30 mV negative to the resting potential (V_R). A possible slight depolarization (5–10 mV) of V_R should have caused a 2–4 times larger decrease in i than actually observed in this experiment. In other experiments (for example, see Fig. 2) i did not change at all. The amplitude and kinetics of Ca²⁺-activated, nonselective channels¹⁸ were also unaffected by the nucleotide, indicating that V_R remained constant (see also ref. 6).

10–25 min after its onset and sampled again 15 min after the washout. With this procedure we found a marked increase in the opening probability, P , of Ca²⁺ channels during exposure to 8-bromocyclic AMP, while the single-channel current, i , changed very little over a voltage range of 40 mV (Fig. 1). Opening probability was calculated from $P = I/Ni$, where I is the mean current found by time-averaging the current flowing through open Ca²⁺ channels during 100-ms depolarizing voltage-clamp steps, and N is the number of Ca²⁺ channels presumably functioning in the patch ($N = 1$ for Fig. 1)⁶. As previously described⁶, Ca²⁺ channels sometimes fail to open during step depolarizations. With 8-bromocyclic AMP we found a significant reduction in the failure of channel opening during repetitive depolarization and, in experiments with more than one channel in the patch, we found on average a higher incidence of multiple conductance levels than in the controls. However, in most experiments the maximum number of conductance levels at each potential remained unchanged, indicating that N was probably not increased by 8-bromocyclic AMP over the tested voltage range.

As expected from Fig. 1, ensemble averages of single-channel traces showed that 8-bromocyclic AMP caused a marked and reversible increase in the mean current amplitude (Fig. 2). Such effects of cyclic AMP and its analogues have been described previously for Ca²⁺ currents in multicellular^{7,8} and single cell^{9,10} cardiac preparations.

* To whom correspondence should be addressed.

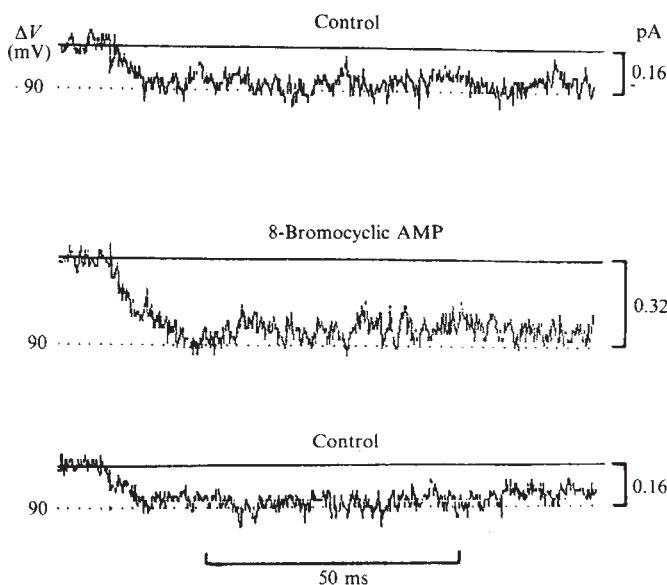


Fig. 2 Overall averages of single Ca^{2+} channel current traces. Mean currents obtained from 219 averaged traces for each condition. Controls: before application (top trace) and after washout (bottom trace) of 8-bromocyclic AMP; centre trace shows the nucleotide effect 15–20 min after its onset. Single-channel currents, i , remained the same (1.12 pA) in all conditions with step depolarizations of 60 mV amplitude, while the average number of open channels increased from 0.14 (controls) to 0.28 in the presence of 8-bromocyclic AMP. The average number of open channels (NP) is found by dividing the average current amplitude, I , by i .

Ca^{2+} channels often open in bursts^{4,6,11} and it is primarily the interval between bursts that is shortened by 8-bromocyclic AMP. This can best be analysed at potential levels where the opening probability of the channels is relatively low. For a statistically reliable analysis we had to sample several hundred current traces in each experimental condition. Two experiments of this kind provided similar results. Representative single Ca^{2+} channel currents, similar to those previously described in detail⁶, are shown in Fig. 3A (upper traces). Quantitative kinetic information on channel openings and closures in control conditions and in the presence of the nucleotide was obtained by plotting open and closed time histograms. The open time histograms (Fig. 3B) could be fitted by single exponentials with time constants, τ_o , of 0.59 ms (control) and 0.66 ms (8-bromocyclic AMP). The slight lengthening of mean open times contributes relatively little to the increase in the overall opening probability of the channels by the nucleotide. A much more dramatic effect can be seen for the closed time histograms (Fig. 4); these histograms were fitted by double exponential curves. While the

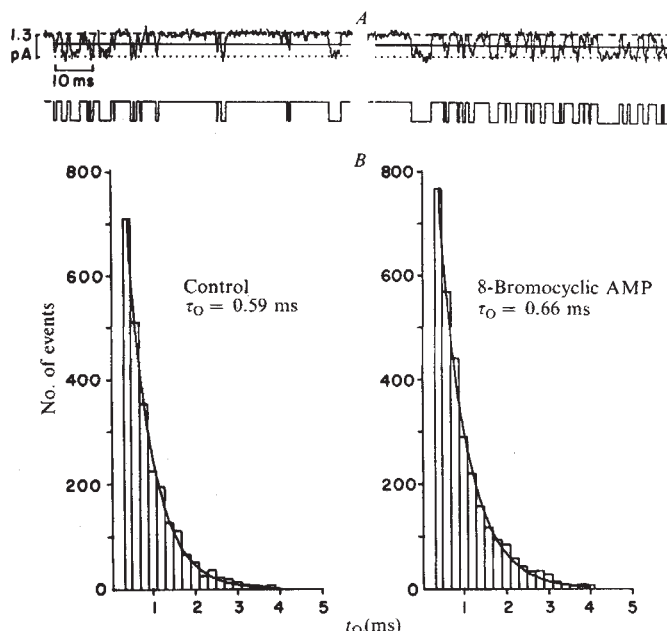
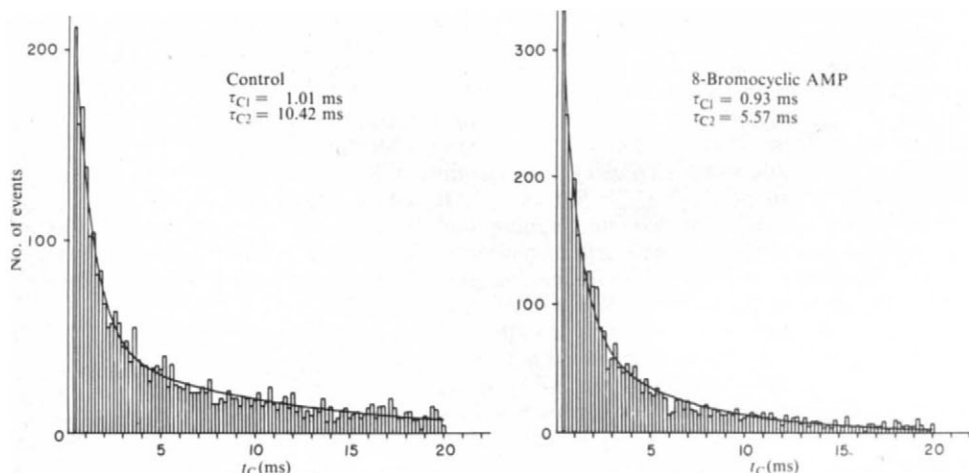


Fig. 3 **A**, Upper traces show single-channel currents and the principles of their analysis⁴. Channel openings occurred during 100-ms voltage clamp steps ($\Delta V = 60$ mV; V_H , resting potential). After subtraction of capacitive transients and leakage current, a base line was fitted to the records by computer (dashed lines) and the average elementary current was estimated by eye (dotted lines); it remained the same in all conditions (1.31 pA). The program detected transitions between conducting and non-conducting states of the channels as crossings of the midline (heavy line) between the two current levels (minimum resolvable interval 0.3 ms). The calculated transition times were displayed by the computer (lower traces) and plotted as histograms. **B**, Histograms of open time distributions were fitted by nonlinear regression analysis (MODFIT¹⁹) with single exponential curves. The time constants (τ_o) corresponding to the mean open times of the channels, were only slightly increased by 8-bromocyclic AMP while the average probability of opening, P , was almost doubled (0.0465, controls; 0.0894, 8-bromocyclic AMP).

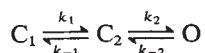
time constant of the fast exponential (τ_{c1}) was only slightly shortened by 8-bromocyclic AMP, that of the slow component (τ_{c2}) was reduced by almost 50%. The short time constants represent the mean durations of the brief closures during burst-like openings, while the longer ones represent the mean durations of the intervals between the bursts. These intervals are clearly much shorter in the presence of the nucleotide, which is the main reason for the increase in P from 0.0465 to 0.0894.

These results can be further analysed on the basis of a kinetic scheme recently proposed for Ca^{2+} channel gating⁴. In this scheme the channel has two closed states, C_1 and C_2 , and one

Fig. 4 Histograms of closed time distributions before (control) and after 8-bromocyclic AMP addition. The histograms were fitted by nonlinear regression analysis with double exponential curves; the respective time constants of the fast and slow components are indicated. The ratios, n , of the integrals of the fast and slow components are 0.510 and 0.705, respectively, and correspond to the mean number of interruptions per burst. Same experiment as in Fig. 3 with identical voltage clamp steps ($\Delta V = 60$ mV).



open state, O, and the reaction between the states can be described as



For such a reaction scheme it can be shown^{4,12} that $\tau_O^{-1} = k_{-2}$ and $\tau_{C_1}^{-1} = k_2 + k_{-1}$. As the average number of interruptions per burst, n , is equal to the ratio of the integrals of the fast and slow components of the closed time histograms⁴ (0.510, control; 0.705, 8-bromocyclic AMP; Fig. 4) and since n is also related to k_2 and k_{-1} by $n = k_2/k_{-1}$ (ref. 4), we can calculate all relevant kinetic parameters of the reaction scheme from the data in Figs 3B and 4. The respective rate constants without and with 8-bromocyclic AMP are k_1 162 s⁻¹ and 211 s⁻¹, k_{-1} 656 s⁻¹ and 631 s⁻¹, k_2 334 s⁻¹ and 444 s⁻¹, and k_{-2} 1,695 s⁻¹ and 1,515 s⁻¹. This result shows that the major effect of 8-bromocyclic AMP on Ca²⁺ channels in the heart is an acceleration of the forward rates leading to channel openings, while the mean open time of the channel ($\tau_O = k_{-2}$) is only slightly prolonged.

We have also obtained very similar results with isoproterenol (refs 2, 6 and A.B.C., J.E. de P. and H.R., unpublished). Both

8-bromocyclic AMP and isoproterenol increase cyclic AMP within the cell, which presumably causes phosphorylation of the Ca²⁺ channel or of a protein closely associated with it¹³⁻¹⁵. The increase in Ca²⁺ current induced by cyclic AMP cannot be explained by a reduction of the free intracellular Ca²⁺ concentration, for example, by activation of the Ca²⁺ pump of the sarcoplasmic reticulum, as the effect persists during perfusion of single cardiac cells with EGTA-buffered solution containing cyclic AMP¹⁰. Moreover, EGTA injection into Purkinje fibres does not prevent the increase in Ca²⁺ current induced by adrenaline¹⁶. Therefore, we suggest that single Ca²⁺ channels in cardiac cells are modulated by cyclic AMP-dependent phosphorylation and that the enhanced Ca²⁺ influx^{13,17} results from a higher opening probability of Ca²⁺ channels in the phosphorylated state. However, our experiments do not rule out an increase in N as an additional effect of Ca²⁺ channel phosphorylation.

This work was supported by the Swiss NSF. We acknowledge the gift of pregnant rats by the Hoffmann-La Roche Animal Farm, Füllinsdorf, and the help of C. Becker in the preparation of cardiac tissue cultures.

Received 21 February; accepted 15 June 1983.

- Hartzell, H. C. *Nature* **291**, 539-544 (1981).
- Reuter, H. *Nature* **301**, 569-574 (1983).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
- Fenwick, E. M., Marty, A. & Neher, E. *J. Physiol., Lond.* **331**, 599-635 (1982).
- Porzig, H., Becker, C. & Reuter, H. *Naunyn-Schmiedeberg's Archs Pharmacol.* **321**, 89-99 (1982).
- Reuter, H., Stevens, C. F., Tsien, R. W. & Yellen, G. *Nature* **297**, 501-504 (1982).
- Tsien, R. W., Giles, W. & Greengard, P. *Nature new Biol.* **240**, 181-183 (1972).
- Reuter, H. *J. Physiol., Lond.* **242**, 429-451 (1974).

- Trautwein, W., Taniguchi, J. & Noma, A. *Pflügers Arch. ges. Physiol.* **392**, 307-314 (1982).
- Irisawa, H. & Kokubun, S. *J. Physiol., Lond.* **338**, 321-337 (1983).
- Brown, A. M., Cammer, H., Kunze, D. L. & Lux, H. D. *Nature* **299**, 156-158 (1982).
- Colquhoun, D. & Hawkes, A. G. *Proc. R. Soc. B211*, 205-235 (1981).
- Reuter, H. A. *Rev. Physiol.* **41**, 413-424 (1979).
- Tsien, R. W. & Siegelbaum, S. in *Physiology of Membrane Disorders* (eds Adreoli, T. E., Hoffman, J. F. & Fanestil, D. D.) 517-538 (Plenum, New York, 1978).
- Manalan, A. S. & Jones, L. R. *J. Biol. Chem.* **257**, 10052-10062 (1983).
- Marban, E. & Tsien, R. W. *J. Physiol., Lond.* **329**, 589-614 (1982).
- Osterrieder, W. et al. *Nature* **298**, 576-578 (1982).
- Colquhoun, D., Neher, E., Reuter, H. & Stevens, C. F. *Nature* **294**, 752-754 (1981).
- McIntosh, J. E. A. & McIntosh, R. P. *Mathematical Modelling and Computers in Endocrinology* (Springer, New York, 1980).

Kluyveromyces lactis killer toxin inhibits adenylate cyclase of sensitive yeast cells

Yuji Sugisaki*, Norio Gunget†, Kenji Sakaguchi†, Makari Yamasaki* & Gakuzo Tamura*

* Department of Agricultural Chemistry, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan
† Mitsubishi-Kasei Institute of Life Science, Minamiooya, Machida-shi, Tokyo 194, Japan

K₁ killer toxin secreted by the K⁺ strain of *Saccharomyces cerevisiae*, has been well characterized^{1,2}. It is a simple protein of molecular weight (MW) 11,470 (ref. 3), encoded by a double-stranded, linear RNA plasmid, called M RNA, of MW 1.1-1.7 × 10⁶ (refs 4-6). It is lethal to sensitive *Saccharomyces cerevisiae* which does not carry M RNA. Leakage of K⁺ and ATP is the first distinct response in sensitive cells⁷⁻⁹, and the toxic action is thought to be due to its action as a protonophore^{10,11} or K⁺ ionophore². Recently, a further killer toxin has been found in *Kluyveromyces lactis* IFO 1267, and it is associated with the presence of the double-stranded linear DNA plasmids, pGK1-1 (MW 5.4 × 10⁶) and pGK1-2 (MW 8.4 × 10⁶)¹². It has been shown, by curing pGK1-1 or deletion mapping, that the structural gene for the killer toxin and immunity-determining gene reside on the smaller plasmid^{13,14}. Moreover, the plasmids could be transferred from *K. lactis* to *S. cerevisiae* by protoplast fusion and protoplast transformation^{13,15}. As the *K. lactis* toxin is encoded by a DNA plasmid and has a relatively wider action spectrum than K₁ killer toxin^{12,13}, the mode of action of the toxin is highly interesting. Here we report that *K. lactis* toxin inhibits adenylate cyclase in sensitive yeast cells and brings about arrest of the cells at the G₁ stage.

The killer toxin was purified 1,560-fold with respect to specific activity from a culture broth of *K. lactis* IFO 1267 by Sephadex G-50 and hydroxyapatite column chromatography (Fig. 1). This preparation was used throughout the present experiment. One batch of the preparation, however, was further purified (200-fold) to 80% purity by chromatofocusing. The purified toxin consisted of two subunits of MW 27,000 (27K)

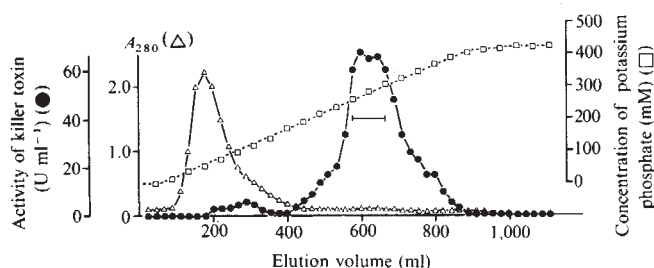


Fig. 1 Purification of *K. lactis* killer toxin on a hydroxyapatite column. *K. lactis* IFO 1267 was grown with agitation in 2 l of a synthetic medium containing 4% glucose, 2% peptone, 1% yeast extract and 0.2% (NH₄)₂SO₄ at 30 °C to the late logarithmic phase. The culture filtrate was concentrated to one-tenth by volume by using hollow fibre H1P10. The concentrated filtrate was applied to a column of Sephadex G-50 (95 cm² × 45 cm) equilibrated with 5 mM potassium phosphate buffer (pH 6.8). The flow-through fraction was applied to a hydroxyapatite column (9 cm² × 15.3 cm) equilibrated with 5 mM potassium phosphate pH 6.8, and absorbed proteins were eluted by a linear gradient concentration of 5-600 mM potassium phosphate buffer pH 6.8. Killer toxin activity was eluted at 250-330 mM potassium phosphate; 110 ml of the peak fractions indicated in the figure were combined and concentrated by a colodion bag to 2 ml (3,500 U ml⁻¹) and dialysed with 0.05 M citric acid-0.1 M Na₂HPO₄ (pH 6.0). The killer toxin concentration of 1 U ml⁻¹ was defined as that which gave a clear zone of 10.3 mm in diameter when a paper disk of 8 mm diameter was moistened with the toxin solution and placed on a standard assay plate at pH 6.0 and 30 °C for 24 h (ref. 20). The killer toxin was purified 1,560-fold with a yield of 33%.

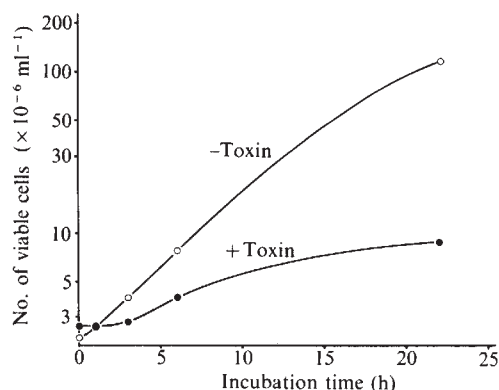


Fig. 2 The effect of *K. lactis* killer toxin on the growth and viability of sensitive *S. cerevisiae*. Logarithmic phase cells of *S. cerevisiae* STX27-1C-1A (α , *ade7*, *met8*, *his7*, *leu2*, *trp*, *gal*) grown in YEPD-CP medium (2% glucose, 2% polypeptone, 1% yeast extract, 0.05 M citric acid–0.1 M Na_2HPO_4 (pH 6.0)) were suspended in fresh YEPD-CP medium at a concentration of $2\text{--}3 \times 10^6$ cells ml^{-1} . *K. lactis* toxin was added to the medium at $1,000 \text{ U ml}^{-1}$ ($\sim 3 \times 10^6$ molecules per cell) and the mixture was kept at 30°C without agitation. An aliquot of sample was withdrawn at certain time intervals, diluted with 0.8% NaCl and spread on YEPD agar medium (2% glucose, 2% polypeptone, 1% yeast extract, 2% agar). Viable counts were made after incubation at 30°C for 2–3 days.

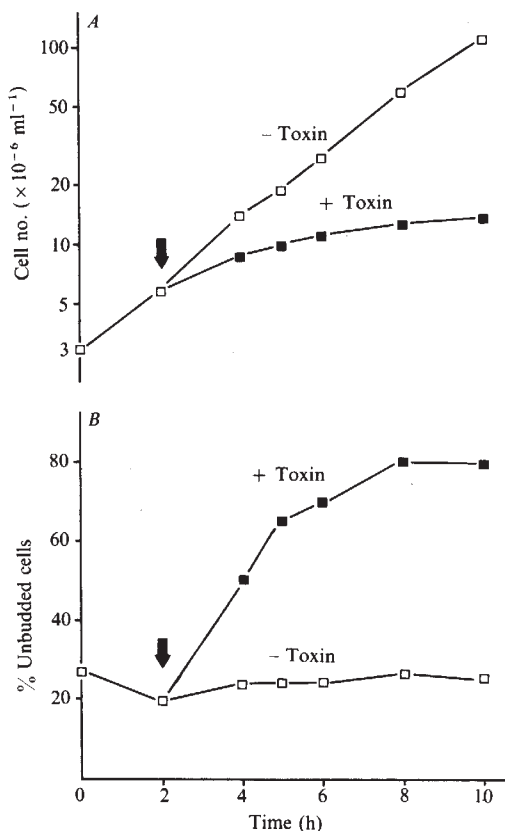


Fig. 3 Effect of *K. lactis* killer toxin on growth (A) and budding (B) of *S. cerevisiae*. *S. cerevisiae* F10D (α , wild type) was grown in YEPD-CP medium to the logarithmic phase (10^7 cells ml^{-1}), then the cells were collected by centrifugation and suspended in fresh YEPD-CP medium to give 3×10^6 cells ml^{-1} and kept at 30°C for 2 h without agitation. *K. lactis* toxin was added to a final concentration of $1,500 \text{ U ml}^{-1}$ and the mixture was incubated at 30°C without agitation. Aliquots were withdrawn at certain time intervals and diluted with 9 vol. of 3% formaldehyde solution in 0.8% NaCl to stop growth. The fixed cells were sonicated (50% pulse) for 20 s to dissociate clumped cells. The total cell number A and unbudded cell fraction B were calculated in a haemocytometer under a microscope. The arrow indicates the time of addition of killer toxin.

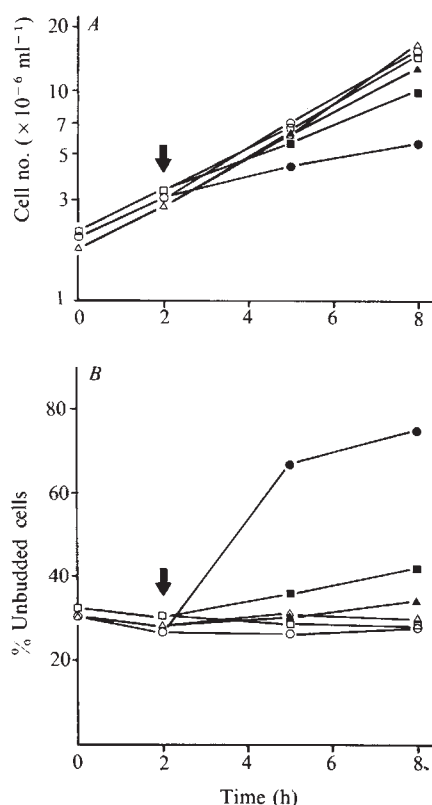


Fig. 4 Inhibitory effect of cyclic AMP and IBMX on action of *K. lactis* toxin on *S. cerevisiae*. Experimental conditions were the same as described in Fig. 3 legend except that 2 mM cyclic AMP or 2 mM cyclic AMP plus 5 mM IBMX was added at zero time, where indicated. Total cell number (A) and unbudded cell fraction (B) were recorded. $\circ$, $\square$, $\triangle$, Without toxin; $\bullet$, $\blacksquare$, $\blacktriangle$, With toxin; $\bullet$, plus toxin only; $\blacksquare$, plus toxin and cyclic AMP; $\blacktriangle$, plus toxin, cyclic AMP and IBMX.

and $>80\text{K}$ (Y.S. *et al.*, in preparation). The number of toxin molecules per treated cell was approximated from the specific activity (100,000 units per mg protein) of the purified preparation assuming a minimum MW of 100,000.

To determine the effect of *K. lactis* killer toxin on the growth and viability of sensitive *S. cerevisiae*, viable counts after treatment with the toxin were examined in a growing culture. The killer toxin greatly suppressed the growth of a sensitive strain of *S. cerevisiae* at excess toxin concentration ($1,000 \text{ U ml}^{-1}$; about 3×10^6 molecules per cell, Fig. 2). The toxin had a static effect on yeast growth to the degree tested in the present experiment. Doubling of the dose of toxin had a similar effect on viability and a dose of 100 U ml^{-1} resulted in less growth inhibition (data not shown). Treatment of sensitive *S. cerevisiae* cells with the toxin did not induce any obvious morphological alteration (data not shown).

Next, we tested the effect of the toxin on the budding of sensitive yeast cells in an unsynchronized liquid culture. About 20% of the cells were always present in the unbudded state in the control culture. After addition of the toxin at $1,500 \text{ U ml}^{-1}$, growth as judged by the total cell count was greatly reduced, and the unbudded cell fraction gradually increased to 80% within ~ 6 h of treatment (Fig. 3). The change in total cell counts coincided with the change in viable counts (see Fig. 2). As buds emerge simultaneously with the initiation of DNA replication¹⁶, the increase in unbudded cell fraction (see Fig. 3) clearly shows that the toxin brought about G_1 arrest in the yeast cells.

Recently, adenylate cyclase-negative mutants of *S. cerevisiae* were found to be arrested at G_1 in the absence of added cyclic AMP¹⁷. This finding recalls the previous observation that yeast sex pheromone α factor inhibits adenylate cyclase and brings about G_1 arrest^{18,19}. We therefore tested the effect of cyclic AMP on the action of *K. lactis* toxin. Cyclic AMP (2 mM) was

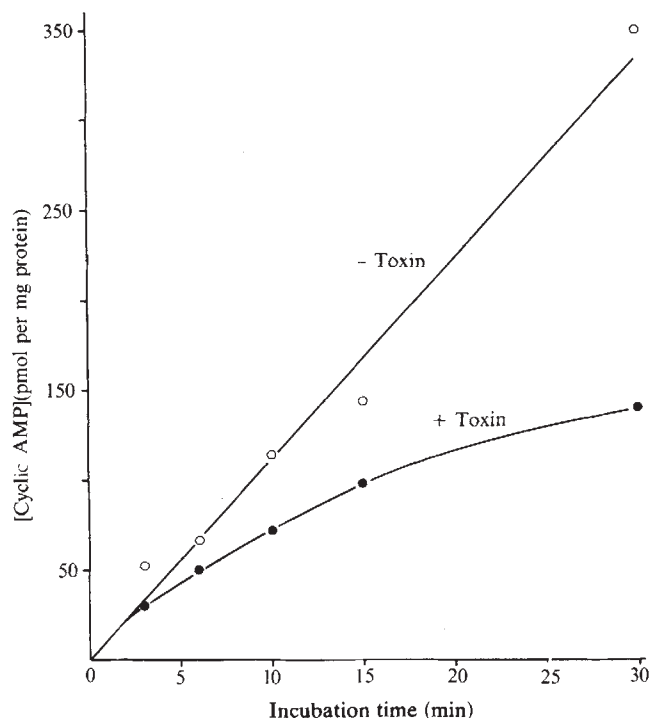


Fig. 5 Effect of *K. lactis* toxin on the activity of adenylate cyclase in a membrane fraction prepared from a sensitive strain of *S. cerevisiae*. The membrane fraction was prepared according to the method of Liao and Thorner¹⁹. *S. cerevisiae* STX27-1C-1A was grown at 30 °C to a density of 10^8 cells ml⁻¹ in 10 ml of YEPD medium. The cells were collected by centrifugation and suspended in 0.5 ml of buffer A (1 M sorbitol, 20 mM potassium phosphate buffer pH 7.5); 50 µl of Zymolyase 6000 (Kirin Brewery Co., Tokyo) solution (1 mg ml⁻¹ in buffer A) were added to the cell suspension, which was kept at 30 °C for 30 min. The resultant protoplasts were washed three times by centrifugation with 0.5 ml of buffer A, and suspended in 5 ml of buffer C (25 mM PIPES-HCl buffer (pH 6.2), 1.1 mM MnCl₂, 1 mM diisopropylfluorophosphate). The diluted protoplast suspension was homogenized in a Potter-Elvehjem homogenizer and centrifuged at 105,000g for 30 min. The pellet was suspended in 1 ml of buffer C and used as a membrane preparation. The protein content of the preparation was measured by the method of Lowry *et al.*²¹. The adenylate cyclase assay mixture contained 25 µl of the membrane preparation, 10 µl of 200 mM PIPES-HCl buffer (pH 6.5), 200 mM creatine phosphate, 1,000 U ml⁻¹ creatine kinase, 5 mM ATP, 50 µl of killer toxin solution (3,500 U ml⁻¹), in a total volume of 100 µl. The reaction mixture was kept at 30 °C for the given time and the reaction was stopped by adding 100 µl of 10% cold trichloroacetic acid (TCA). The mixture was centrifuged at 12,000 r.p.m. for 3 min and the supernatant was collected. The precipitate was suspended in 50 µl of 5% TCA and centrifuged; the resultant supernatant was combined with that from the first centrifugation. TCA was removed by repeated extraction with water-saturated ether. Cyclic AMP content was then estimated by a conventional radioimmunoassay method by using a cyclic AMP assay kit (Yamasa Shoyu Co., Choshi, Japan)²².

added to the reaction system shown in Fig. 3, and the growth and size of the unbudded cell fraction were monitored. Figure 4 shows that 2 mM cyclic AMP effectively blocked the inhibitory action of the toxin on growth and budding of the yeast cells. Simultaneous addition of 5 mM 3-isobutyl-1-methylxanthine (IBMX), a potent inhibitor of yeast cyclic AMP phosphodiesterase¹⁹, to 2 mM cyclic AMP enhanced the effect of cyclic AMP. Cyclic GMP, 5'-AMP and adenosine at 2 mM could not substitute for cyclic AMP (data not shown). These results suggest that *K. lactis* toxin inhibits the activity of adenylate cyclase in sensitive cells. We tested the effect of the toxin on the activity of adenylate cyclase in a crude membrane preparation of sensitive *S. cerevisiae* (Fig. 5). The toxin inhibited adenylate cyclase activity at a comparable concentration (5 ×

10^5 molecules per cell) to that adopted in the *in vivo* system. A lag time of ~30 min was observed before complete inhibition of adenylate cyclase activity. This may represent the time required for putative activation of the toxin.

Thus, *K. lactis* toxin may bring about G₁ arrest by inhibiting adenylate cyclase activity. Hence the action of the toxin was yeast-static, although other modes of action are possible. The action of *K. lactis* toxin proposed here is in sharp contrast to that of double-stranded linear RNA-encoded K₁ killer toxin, which is lethal due to its action as a K⁺ ionophore or protonophore^{2,10,11}. In fact, *K. lactis* did not exhibit the ATP leakage effect demonstrated for K₁ killer toxin (data not shown).

The identification of a putative toxin receptor, which may be adenylate cyclase itself or another molecule and determines the sensitivity of yeasts to the toxin, and the elucidation of the fine inhibitory mechanism of adenylate cyclase await further study.

We thank Drs K. Matsumoto (University of Tottori) and I. Uno (University of Tokyo) for supplying yeast strains and for their valuable suggestions and discussion. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

Received 29 March; accepted 27 May 1983.

- Wickner, R. B. *The Molecular Biology of the Yeast Saccharomyces* (eds Strathern, J. N., Jones, E. W. & Broach, J. R.) 415-444 (Cold Spring Harbor Laboratory, New York, 1981).
- Bussey, H. *Adv. microbiol. Physiol.* **22**, 93-122 (1981).
- Palfrey, G. E. & Bussey, H. *Eur. J. Biochem.* **93**, 487-493 (1979).
- Fink, G. R. & Styles, C. A. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2846-2849 (1972).
- Wickner, R. B. *J. Bact.* **117**, 1356-1357 (1974).
- Bostian, K. A., Hopper, J. E., Rogers, D. T. & Tipper, D. J. *Cell* **19**, 403-414 (1980).
- Bussey, H. & Sherman, D. *Biochim. biophys. Acta* **298**, 868-875 (1973).
- Kotani, H., Shinmyo, A. & Enatsu, T. *J. Bact.* **129**, 640-650 (1977).
- Skipper, N. & Bussey, H. *J. Bact.* **129**, 668-677 (1977).
- de la Peña, P., Barros, F., Gascón, S., Ramos, S. & Lazo, P. S. *Biochem. biophys. Res. Commun.* **96**, 544-550 (1980).
- de la Peña, P., Barros, F., Gascón, S., Lazo, P. S. & Ramos, S. *J. biol. Chem.* **256**, 10420-10425 (1981).
- Gunge, N., Tamaru, A., Ozawa, F. & Sakaguchi, K. *J. Bact.* **145**, 382-390 (1981).
- Gunge, N. & Sakaguchi, K. *J. Bact.* **147**, 155-160 (1981).
- Niwa, O., Sakaguchi, K. & Gunge, N. *J. Bact.* **148**, 988-990 (1981).
- Gunge, N., Murata, K. & Sakaguchi, K. *J. Bact.* **151**, 462-464 (1982).
- Hartwell, L. H. *Bact. Rev.* **38**, 164-198 (1974).
- Matsumoto, K., Uno, I., Oshima, Y. & Ishikawa, T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2355-2359 (1982).
- Liao, H. & Thorner, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1898-1902 (1980).
- Liao, H. H. & Thorner, J. *J. Bact.* **148**, 919-925 (1981).
- Woods, D. R. & Bevan, E. A. *J. gen. Microbiol.* **51**, 115-126 (1968).
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265-275 (1951).
- Homma, M., Satoh, T., Takezawa, J. & Ui, M. *Biochem. Med.* **18**, 257-273 (1977).

Inducible repair of oxidative DNA damage in *Escherichia coli*

Bruce Demple

Imperial Cancer Research Fund, Mill Hill Laboratories, Burtonhole Lane, London NW7 1AD, UK

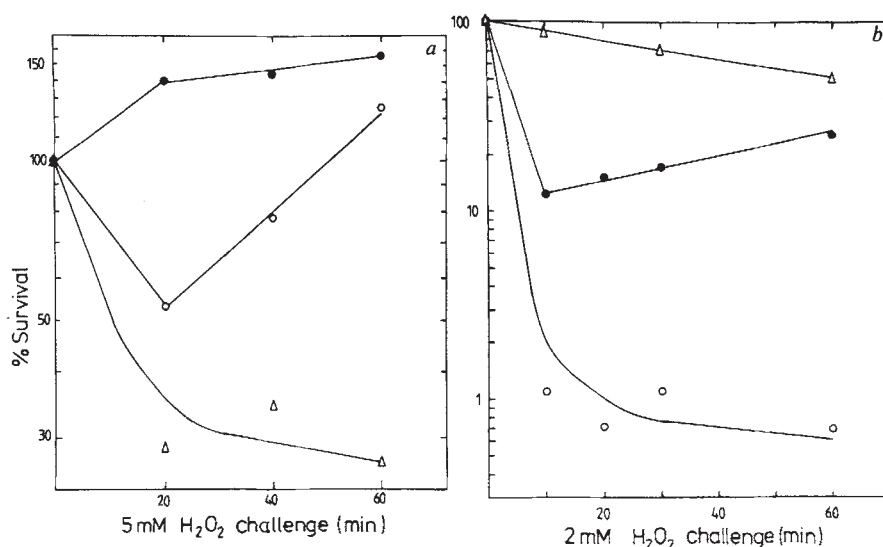
James Halbrook*

Department of Biochemistry, University of California, Berkeley, California 94720, USA

Hydrogen peroxide is lethal to many cell types, including the bacterium *Escherichia coli*^{1,2}. Peroxides yield transient radical species that can damage DNA³ and cause mutations⁴. Such partially reduced oxygen species are occasionally released during cellular respiration² and are generated by lethal and mutagenic ionizing radiation⁵. Because cells live in an environment where the threat of oxidative DNA damage is continual, cellular mechanisms may have evolved to avoid and repair this damage. Enzymes are known which evidently perform these functions^{2,3}. We report here that resistance to hydrogen peroxide toxicity can be induced in *E. coli*, that this novel induction is specific and occurs, in part, at the level of DNA repair.

* Present address: Department of Biological Chemistry, University of California School of Medicine, Los Angeles, California 90024, USA.

Fig. 1 Induction of bacterial resistance to H_2O_2 . **A**, Induced resistance in *E. coli* AB1157. Strain AB1157 (ref. 18) was grown at 37 °C in K medium (M9 medium²⁴ containing 1% glucose, 1% casamino acids, $1 \mu\text{g ml}^{-1}$ thiamine-HCl, 1 mM $\text{MgSO}_4 \cdot 7H_2O$ and 0.1 mM CaCl_2) to a density of 5×10^7 cells ml^{-1} . Aliquots (3 ml) were removed to separate flasks, and 3 mM H_2O_2 (Fluka; stock concentration 30%) was added to give 0 (Δ), 10 μM ($\circ$) or 30 μM ($\bullet$) final concentrations. After 30 min shaking at 37 °C, aliquots were removed to titre for colony-forming ability, 1 M H_2O_2 was added to each culture to a final concentration of 5 mM, and shaking at 37 °C continued. Aliquots were removed as indicated and diluted 100-fold to assay colony-forming ability. All dilutions for plating were in M9 buffer, and samples were plated on LB agar²⁴ and grown at 37 °C. In the experiment shown here, only 0.1% casamino acids were provided in the medium, and stationary phase was $\sim 2 \times 10^8$ bacteria ml^{-1} . **B**, Induced resistance in *recA*⁻ *E. coli*. Cultures of JC4583 (*recA*⁺; ref. 18) and JC4588 (*recA*⁻; ref. 18) in K medium were grown at 37 °C to a density of 6×10^7 ml^{-1} and divided into equal aliquots. H_2O_2 (to give 15 μM) was added to one of the *recA*⁻ cultures and shaking continued for 40 min, then a second 15- μM aliquot of H_2O_2 was added to the same culture. After an additional 40 min growth at 37 °C, samples were removed to titre for colony formation and H_2O_2 added to give 2 mM in each culture. Colony-forming ability was assayed as described above, except that M9 buffer containing 1% glucose and 10 mM $\text{MgSO}_4 \cdot H_2O$ was used as the diluent. Δ , JC4588 *recA*⁺, no pretreatment; $\circ$, JC4588 *recA*⁻, no pretreatment; $\bullet$, JC4588 *recA*⁻, pretreated.



Exponentially growing *E. coli* K-12 were killed rapidly when 5 mM H_2O_2 was introduced into the growth medium (Fig. 1A). As expected, peroxide concentrations of 0.2–1 mM were less toxic, but those conditions gave rise to unusual 'V-shaped' survival curves with time. One explanation for initial killing followed by renewed growth is that the surviving bacteria have been able to induce resistance to the damaging effects of H_2O_2 . Indeed, prior treatment of growing cells with 10 μM H_2O_2 reduced the initial toxicity of 5 mM peroxide (Fig. 1A). The survivors after 20 min even began growing at the pretreatment rate, generating a V-shaped survival curve reminiscent of the profiles we observed for non-pretreated *E. coli* challenged with moderate H_2O_2 concentrations. Bacteria previously exposed to 30 μM H_2O_2 for 30 min survived the 5 mM peroxide challenge completely and continued to grow at the normal rate to stationary phase. Such pretreated cultures also showed considerably enhanced survival in still higher peroxide concentrations (up to 25 mM). A consistent 10–20-fold increase in the resistance of pretreated *E. coli* to 20 mM H_2O_2 has been observed in experiments with several *E. coli* K-12 and B/r strains, and in different growth media. Cells pretreated in the presence of the protein synthesis inhibitor, chloramphenicol, were not refractory to hydrogen peroxide toxicity. The degree of induced H_2O_2 resistance increased with the pretreatment level used from 5 μM to 50 μM H_2O_2 ; peroxide concentrations of $\geq 100 \mu\text{M}$ were initially bacteriostatic or lethal.

Induced resistance to hydrogen peroxide toxicity is not part of any previously described inducible resistance to DNA damage. *E. coli* pretreated with peroxide were just as sensitive to killing by the alkylating carcinogen, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine, as were their counterparts that received no pretreatment; thus the induction described here is different from the adaptive response to alkylation damage⁶. Induction of peroxide resistance also failed to increase the survival of *E. coli* after UV light treatment, distinguishing this phenomenon from the *recA*⁺-dependent SOS response⁷. Moreover, a pronounced resistance to H_2O_2 was induced in a *recA*⁻ mutant (Fig. 1B), although a gentler approach was needed due to the previously recognized peroxide sensitivity of *recA*⁻ strains^{8,9}. The induction we describe is also distinct from the recently described 2-aminopurine-induced DNA repair which, like SOS, acts on UV-induced damage¹⁰.

Other workers have reported that the protective enzymes superoxide dismutase¹¹ (catalysing the breakdown of the superoxide radical O_2^-) and catalase¹² (which destroys H_2O_2) are

inducible. The level of superoxide dismutase in bacterial extracts was unchanged following the induction of peroxide resistance (Table 1). A moderate increase (2–4-fold) in catalase activity, however, was obtained by H_2O_2 pretreatment, but this was variable and did not correlate with the degree of H_2O_2 resistance induced in the bacteria, which was the same in all the experiments reported in Table 1. Others have shown⁹ that,

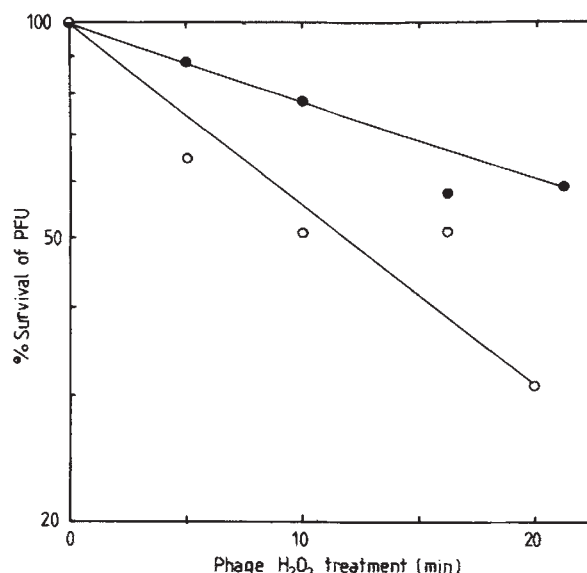


Fig. 2 Host cell reactivation of oxidized bacteriophage P1. A stock of bacteriophage P1 *cml clt* 100 (10^9 plaque-forming units (PFU) ml^{-1}) was prepared from a lysogen according to Rosner¹³ and diluted for treatment in M9 buffer containing 10 mM $\text{MgSO}_4 \cdot 7H_2O$. Equal aliquots of phage and 2 mM H_2O_2 , 10 μM CuSO_4 were mixed at room temperature (22 °C). At the indicated times, bovine catalase (74,600 units per mg; Worthington-Millipore) was added to 15 $\mu\text{g ml}^{-1}$, the incubation continued for 5 min and the samples placed on ice. For the '0 min' treatment, peroxide and catalase were mixed 5 min before the addition of the virus suspension. Catalase and H_2O_2 were diluted in the same buffer as the phage. For titrating, AB1157 were grown to 2×10^8 cells ml^{-1} in K medium at 37 °C and either pretreated or not with 30 μM H_2O_2 for 35 min. Treated phage (5 or 20 μl) were adsorbed to 0.2 ml of untreated ($\circ$), or H_2O_2 -pretreated ($\bullet$) bacteria for 10 min before plating on LB agar to assay for infectious centres at 42 °C.

in general, superoxide dismutase and catalase activities are poor indicators of sensitivity to H_2O_2 .

These results suggest that DNA repair activities are induced during H_2O_2 pretreatment. This hypothesis was tested by examining the ability of cells to repair (reactivate) incoming damaged DNA. Samples of a bacteriophage P1 suspension were incubated *in vitro* with H_2O_2 and Cu^{2+} (to catalyse H_2O_2 breakdown) and plated onto naive or H_2O_2 -induced *E. coli*. Pretreatment of the host bacteria significantly increased the production of progeny virus as measured by plaque formation (Fig. 2). H_2O_2 -damaged P1 (bearing a gene for chloramphenicol resistance¹³) also lysogenized the pretreated cells more efficiently, as monitored by the production of chloramphenicol-resistant lysogens (not shown). The ability of undamaged P1 to lyse or lysogenize *E. coli* was not changed by pretreatment of the

Table 1 Effect of H_2O_2 pretreatment of *E. coli* on scavenging enzyme levels

Source of extract	Experiment	Specific activity (units mg^{-1})	
		Superoxide dismutase	Catalase
No pretreatment	I	4.3	3.2
	II	6.3	4.2
H_2O_2 -pretreated	I	5.8	3.7
	II	4.7	14.4
	III	NT	6.6

Cultures of AB1157 in K medium were grown to 2×10^8 cells ml^{-1} and pretreated as described in Fig. 1 legend. Extracts were prepared according to Wickner *et al.*²⁰ and assayed immediately for catalase activity²¹. Superoxide dismutase activity²² was determined with either fresh or frozen extracts, and one cycle of freeze-thawing did not affect the activity. Protein concentrations were measured by the method of Lowry *et al.*²³. NT, Not tested.

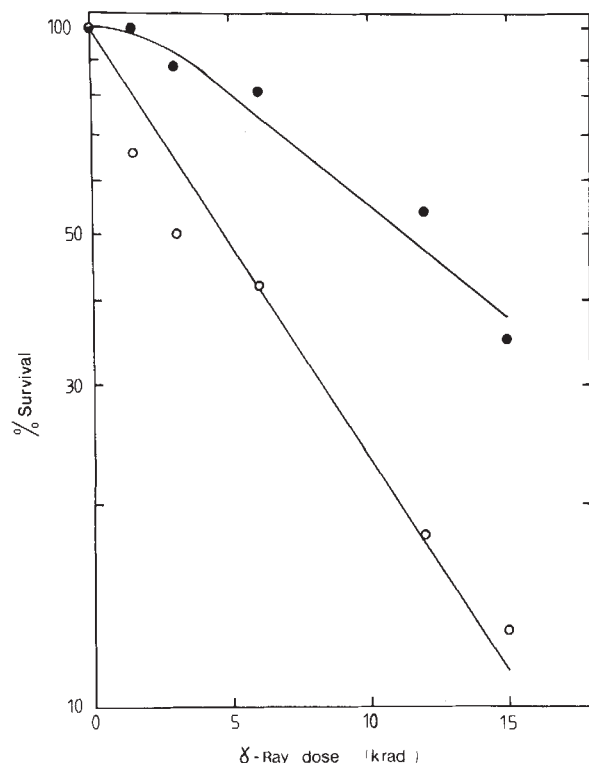


Fig. 3 Induction of γ -ray resistance by H_2O_2 . Strain AB1157 was grown to 10^8 cells ml^{-1} in K medium and pretreated (●) or not (○) with $30 \mu M$ H_2O_2 for 40 min. Catalase was added to $2 \mu g$ ml^{-1} in both cultures for the last 10 min of pretreatment, and the cultures placed on ice. Aliquots (1 ml) in sterile bottles were irradiated in parallel at a dose rate of $6,800$ rad min^{-1} from a ^{137}Cs source. Surviving bacteria were assayed as described in Fig. 1 legend.

bacteria with peroxide. These results demonstrate directly that repair of H_2O_2 damage in DNA is specifically induced. Although the magnitude of this effect on phage survival may not seem large (a 2–3-fold increase in reactivation has been observed in several experiments), it is similar to the enhanced reactivation of alkylated virus achieved by adaptation of the host cells with nitrosoguanidine¹⁴.

As H_2O_2 and ionizing radiation yield similar reactive species, we tested the effect of H_2O_2 pretreatment on the resistance of *E. coli* to γ rays. Prior treatment with peroxide doubled the survival rate of strain AB1157 (Fig. 3). Similar gains in γ -ray resistance have been measured in experiments with doses up to 50 krad, irrespective of the presence of added catalase in the medium (see Fig. 3 legend). However, the enhancement of survival seems greatest for radiation doses ≤ 10 krad. We conclude that peroxide induces the repair of lesions produced both by itself and by ionizing radiation. In this regard, *E. coli* mutants selected for their resistance to γ rays are reported to be resistant to H_2O_2 also¹⁵.

Among the enzymes in *E. coli* known to act specifically on oxidized DNA are the *N*-glycosylases for thymine glycol¹⁶ and urea¹⁷. Both were present at the same level in bacterial extracts with or without H_2O_2 pretreatment. Exonuclease III is implicated in the repair of oxidized DNA because *E. coli* which lack the enzyme are highly sensitive to H_2O_2 (ref. 18). However, resistance to H_2O_2 lethality could be induced in such a strain, and the activity of exonuclease III in extracts of wild-type *E. coli* was unaffected by peroxide pretreatment. Present experiments are directed at examining the repair of other peroxide-induced oxidative DNA lesions in *E. coli*, but functions mobilized to delay replication of oxidized DNA could also be involved.

The phenomenon described here is distinct from, but has similarities to, the adaptive response to alkylating agents^{6,19}: both repair pathways are induced by non-toxic levels of the respective agents and increase the cellular capacity to correct a specific class of lethal damage to DNA. Alkylation adaptation also leads to a striking ability of pretreated bacteria to correct important premutagenic lesions. This is not the case with the SOS response, which includes a considerable increase in mutagenesis by many agents⁷. Preliminary experiments indicate that peroxide-mediated mutagenesis is not increased in H_2O_2 -pretreated AB1157. Although H_2O_2 is only moderately mutagenic⁴, experiments are under way to test whether peroxide pretreatment reduces H_2O_2 mutagenesis. At present it seems that non-toxic levels of hydrogen peroxide can elicit another kind of 'adaptive response' in *E. coli*, one directed at oxidative damage.

We thank Stuart Linn and Tomas Lindahl for support and helpful discussions. B.D. is a fellow of the Helen Hay Whitney Foundation.

Received 29 March; accepted 8 June 1983.

- McCormick, J. P., Fischer, J. R., Pachlatko, J. P. & Eisenstark, A. *Science* **191**, 468–469 (1976).
- Fridovich, I. *Science* **201**, 875–880 (1978).
- Demple, B. & Linn, S. *Nucleic Acids Res.* **10**, 3781–3789 (1982).
- Levin, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7445–7449 (1982).
- Repine, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 1001–1003 (1981).
- Samson, L. & Cairns, J. *Nature* **267**, 281–283 (1977).
- Witkin, E. *Bact. Rev.* **40**, 869–907 (1976).
- Ananthaswamy, H. N. & Eisenstark, A. *J. Bact.* **130**, 187–191 (1977).
- Carlsson, J. & Carpenter, V. S. *J. Bact.* **142**, 319–321 (1980).
- Maenhaut-Michel, G. & Caillet-Fauquet, P. *Molec. gen. Genet.* **188**, 143–148 (1982).
- Gregory, E. M. & Fridovich, I. *J. Bact.* **114**, 543–548 (1973).
- Richter, H. E. & Loewen, P. C. *Biochem. biophys. Res. Commun.* **100**, 1039–1046 (1981).
- Rosner, J. L. *Virology* **48**, 679–689 (1972).
- Evensen, G. & Seeberg, E. *Nature* **296**, 773–775 (1982).
- Bresler, S. E., Verbenko, V. N. & Kalinin, V. L. *Genetika* **16**, 1753–1763 (1980).
- Demple, B. & Linn, S. *Nature* **287**, 203–208 (1980).
- Breimer, L. & Lindahl, T. *Nucleic Acids Res.* **8**, 6199–6211 (1980).
- Demple, B., Halbrook, J. & Linn, S. *J. Bact.* **153**, 1079–1082 (1983).
- Jeggo, P., Defais, M., Samson, L. & Schendel, P. *Molec. gen. Genet.* **157**, 1–9 (1977).
- Wickner, W., Schekman, R., Geider, K. & Kornberg, A. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1764–1767 (1973).
- Beers, R. F. Jr & Sizer, I. W. *J. biol. Chem.* **195**, 133–140 (1952).
- Heikkila, R. E. & Cabbat, F. *Analyt. Biochem.* **75**, 356–362 (1976).
- Lowry, O., Rosebrough, N., Farr, A. & Randall, R. *J. biol. Chem.* **193**, 265–275 (1951).
- Miller, J. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).

Art and archaeology

Peter Warren

Thera: Pompeii of the Ancient Aegean. By Christos G. Doumas.
Thames & Hudson: 1983. Pp.168. £16, \$29.95.

THE excavations of Spyridon Marinatos from 1967 to 1974 at the site of Akrotiri on the Greek island of Thera (Santorini) revealed a Bronze Age (c. 1500 BC) town of noble mansions, brilliantly decorated with wall paintings, the whole preserved under huge deposits of pumice from the eruption of the Thera volcano. A general account of the discoveries, well illustrated with technical drawings as well as art photographs, can thus hardly fail to succeed, especially when it is written by the archaeologist who has directed operations there since Marinatos' death on site in 1974.

Professor Doumas brings to his writing command of the scores of specialist publications and two international congresses already devoted to the site and the volcano, and a unique authority derived from a director's responsibility for understanding a site as a whole and the interdependence of all its finds. His book is particularly good in dealing with a range of detailed topics. Friedrich's discoveries of the palaeovegetation (back to 37,000 BP) are carefully summarized, there are excellent accounts of nineteenth-century excavations and evaluation of their finds, and there is much that will interest on the technical aspects of excavation and site conservation. Of most value, however, is the lengthy description and careful analysis of the original positions of the many different and always remarkable paintings on the walls. Years of study by Doumas and his colleagues have in some cases enabled surer placings of paintings than Marinatos was able to determine initially. Only from such knowledge can sound iconographical interpretation proceed.

Such a wealth of material has to be integrated with what is already known from the prehistoric Aegean. Discussion is taking place at all levels, in studies of artefacts (especially the ceramics and stone vessels), architecture and iconography. These lead to difficult

but stimulating questions of explaining Akrotiri's wealth within Aegean economic and social models. Her destruction too is a major focus of interest. Three points may be briefly touched on.

First, Doumas is inclined to accept religious functions for the wall paintings of the lilies and swallows, of the bare-breasted women in the House of the Ladies, and of the women gathering saffron and presenting it to a seated figure. The find-contexts of the first and third paintings give additional support to the interpretation. In the case of the second, the bare-breasted ladies bending forward, there is (*pace* p.82) good comparative iconographical evidence that the large papyrus plants, painted nearby, were also held sacred. The boxing boys in House Beta may also have been engaged in

ritual activity, for one of them wears the Minoan sacred knot in his waist, like a bronze figure from a Cretan sanctuary at Kato Syme.

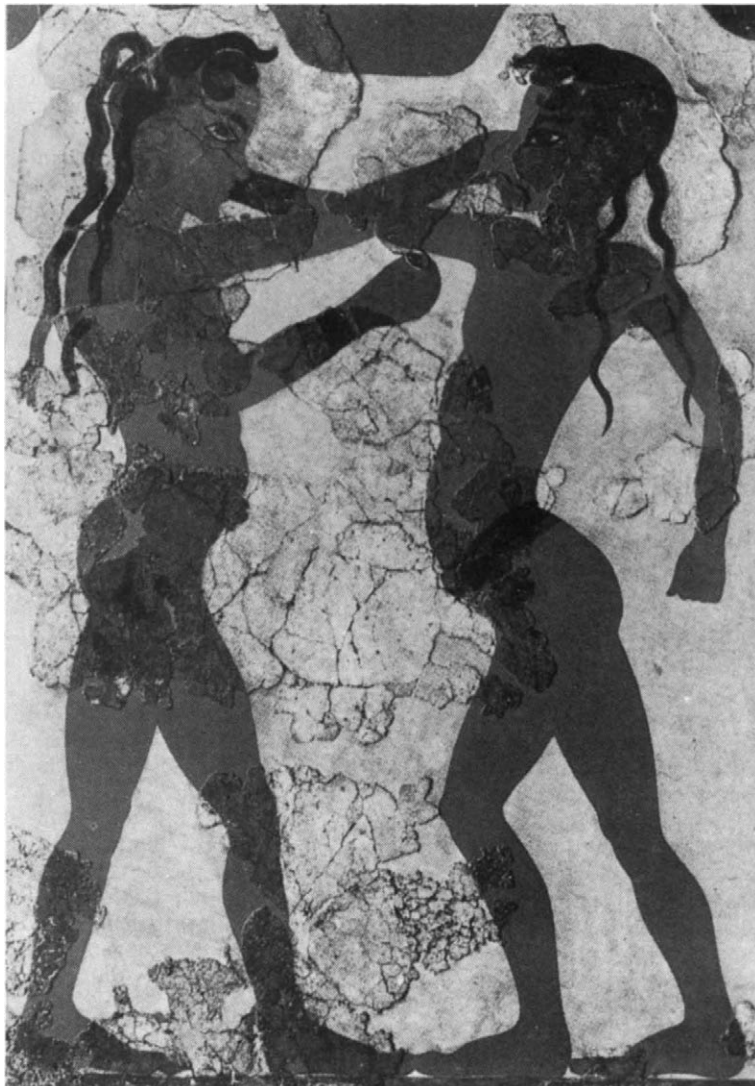
Second, the view consistently taken that Cretan art was restrained in spirit and zonal in form, in contrast to the vitality of Thera and Cycladic art, is scarcely tenable. The fondness for painting swallows and other birds in flight may well be a Thera creation. On the other hand the wind-blown style of plant painting on some Thera vases may well derive from east Crete. As for wall paintings, the vitality of colouring, freedom of surface and empathy with nature in the paintings of buildings such as the House of the Frescoes at Knossos, Aghia Triadha and the villa at Amnisos not only provide parallels for Thera frescoes, but were in all probability the prototypical tradition for them.

Finally, Doumas takes the date of the Thera eruption to be Late Minoan I A (c. 1500 BC), on the grounds that it was the cause of the destruction of the Akrotiri site. He and others who share this view may be right. But it must be noted that Doumas himself stresses the extent of clearance operations during an interval between the

destruction of the site and the last and greatest ash fall; he elects to give the interval phases a duration of up to more than two years (pp. 134–135); the matter appears open. Some of the comparative evidence cited for the LM I A date is quite insufficient (Rhodes) or not published stratigraphically (Zakros). Contrary to what is stated Knossos, like all the other Cretan sites, suffered a major destruction in LM I B, c. 1450 BC (the alternative possibility for the Thera eruption). So far too was Crete from decline in the LM I B phase (1500–1450 BC), i.e. after the eruption on Doumas's view, that this period stands in fact as the zenith of Minoan civilization.

These points of discussion serve to illustrate what an immensely important and productive site Akrotiri is and will continue to be. This book will rightly be a large success and one may anticipate future editions in which Professor Doumas incorporates his further discoveries and valued analyses. □

Peter Warren, Professor of Ancient History and Classical Archaeology at the University of Bristol, has just finished a five-year excavation at Knossos.



Theran art — the boxing boys fresco in House Beta

Two sides to the brain

John J. Sidtis

Human Cerebral Asymmetry.

By John L. Bradshaw and Norman C. Nettleton.

Prentice-Hall: 1983. Pp.335. \$24.95, £22.45.

OVER the past twenty-five years, a feature of neuropsychology has been an enduring interest in the functional differentiation of the human cerebral hemispheres. Beginning with the early attempts to model a functional architecture of the brain based on clinical observations, experimental studies have gone on to delineate component processes as the province of one cerebral hemisphere or the other.

Such studies led to new general descriptions of the functional properties of the two hemispheres. Whereas the neurological distinctions of major versus minor, and dominant versus non-dominant were based on the observation that language is typically represented in a single hemisphere, usually the left, new distinctions such as verbal versus non-verbal and analytic versus holistic acknowledge the fact that the non-language hemisphere is dominant for some non-linguistic functions. Hemispheric differences have also attracted popular interest, frequently because of wild promises that hidden skills, untapped creativity or inner peace could be obtained simply by unleashing one's non-verbal hemisphere.

Given the attention paid to the differences in the two sides of the brain, it is quite surprising that until recently there were no introductory books on the topic. There are now several, with Bradshaw and Nettleton's *Human Cerebral Asymmetry* providing the broadest perspective. The authors cover the major topics such as how asymmetrical brain function can be observed through perceptual asymmetries in vision, audition and touch; the role of development in cerebral asymmetry; the possible genetic determination of cerebral asymmetry; and the evidence that there are sex differences in the degree of cerebral asymmetry. In addition, there is careful and balanced consideration of other topics, such as functional brain asymmetries in animals and morphological brain asymmetries in humans that are frequently presented in secondary sources in a way that exaggerates their relevance to functional cerebral asymmetry in humans.

The authors are at their best in discussing the wide range of studies on human subjects, the population they have been most involved with in their own research. Work on people with neurological damage is covered only nominally, the emphasis being on language dysfunction. The chapter on that topic begins with some

reasonable caveats regarding the interpretation of neurological data, then proceeds to a section on clinical techniques dealing with the establishment of hemispheric dominance for language. A useful addition here, and to the book as a whole, would have been an account of the most important clinical technique, the neurological examination. Neuropsychology has grown as much from clinical neurology as it has from experimental psychology, and in spite of the uncertainties inherent in studying clinical populations, the nature of the enterprise means that theories about brain-behaviour relationships derived from the study of normal subjects must eventually be validated in neurological populations.

The book's weakest point is in the coverage of the work on patients who have undergone surgical separation of the cerebral hemispheres for the control of

epilepsy. Unfortunately, Bradshaw and Nettleton have limited themselves in many instances to the results obtained from two patients in a single surgical series. Moreover, a long section is given over to an account of several testing techniques that have not contributed substantially to understanding either the split-brain or normal cerebral asymmetry.

In general, though, Bradshaw and Nettleton have written an excellent introduction to the wide range of subjects that constitute the study of how function is represented in the two hemispheres of the brain. *Human Cerebral Asymmetry* is commendable for its style, balance and thoughtful presentation. □

John J. Sidtis is an Assistant Professor of Psychology in Neurology at Cornell University Medical College, and a Professional Associate at the New York Hospital.

Palaeontology by the telephone book

D.V. Ager

Fossil Invertebrates.

By U. Lehmann and G. Hillmer.

Translated by Janine Lettau.

Cambridge University Press: 1983.

Pp.355. Hbk £20, \$39.50;

pbk £7.50, \$13.95.

"FACTS are essential!" said Michael Stapleton in his recent volume on English literature, but "a guide who never makes a comment makes a dull companion". That's how I feel about this book. I know that some of my palaeontology is now a little taxonomically dated and I'm not too sure about the Lychniskida and the Caudofoveata, but I am surprised to see memories of the past such as the Class Lamelli-branchia (to say nothing of quaint old terms such as "Dogger" and "Malm").

This is a thorough work of German scholarship; in fact it out-Zittels Zittel, though at much shorter length. One may ask if it is needed these days with R.C. Moore's vast *Treatise on Invertebrate Paleontology* and the fat American, French and Russian volumes that continue to pour upon us, but it is in many ways more up to date and the best treatise in the world inevitably become dated as soon as it is published. I have not seen the original German version, but the translation seems competent and not unduly stiff, without adapting itself to many of the usages of English-speaking palaeontology and stratigraphy.

Although it is called *Fossil Invertebrates*, the book goes beyond the animal kingdom in the usual sense, at least in the microscopic field, by including several groups of algae, bacteria, the fungi (certainly not one of the more obvious

fossil groups) and even the Acritarchs and the Chitinozoa. This makes it more useful if a little confusing, as does the placing of some of the groups (the conodonts, for example, come between the brachiopods and the echinoderms).

Apart from a brief section on the origin of life, there is very little on evolution, or palaeoecology, or functional morphology, or geographical distribution or any of a dozen other subjects that many of us find so fascinating. No doubt that makes it the more scientific in a purely descriptive sense; it also — regrettably — makes it extremely dull. This is wholeheartedly a reference book in systematic palaeontology, not a book to read. It is not large enough to tackle genera, like Zittel or Moore, so it has to concentrate higher up the hierarchical scale. Many of us will probably think that this makes it less scientific rather than more so, since the only biological reality (difficult though it may be) is the species, and the rest are man-made fictions. I would argue, for example, (though perhaps it is a personal idiosyncrasy) about the validity of the Spiriferida as a natural division of the organic world. Other specialists would doubtless disagree about other groups.

What is this book for? Who reads telephone directories? At least it will be useful to me in understanding the examination questions of some of my younger colleagues, but where is the excitement of the subject that I found so long ago in Hawkins and Swinnerton, and so much later in Raup and Stanley? I write this review in hospital so I crave forgiveness for any drug-induced sourness that may have crept in; but this really is not a very inspiring palaeontological text and certainly not one that I would press upon either student or researcher. □

Derek Ager is Professor and Head of the Department of Geology at University College, Swansea.

Place for all of the mammals

R.J.G. Savage

Mammalian Paleofaunas of the World.

By Donald E. Savage and Donald E. Russell.

Addison-Wesley: 1983. Pp.432. \$79.95, £55.20.

THIS book is set fair to become the most quoted work in mammalian evolution, palaeogeography and biostratigraphy. The authors set out to list every mammal genus and species with a fossil record, a task as unambiguous as it is formidable — more than 3,000 genera and at least twice that number of species, over a 200-million-year time-span, and spread across six continents (the discovery of a fossil marsupial in Antarctica came too late for inclusion).

However, this daunting task has been made lighter by the recent publication of major treatises on Mesozoic and Pleistocene mammals*. By dealing with these periods relatively briefly, the authors are left with the Tertiary faunas, one-third of the time span and 90% of the known taxa. North America and Europe get complete coverage though other continents are treated less fully. Thus South American faunas are mostly quoted at generic level; African faunas are not as up-to-date as they might be; Asiatic faunas are rather patchily covered and the Russian examples are especially poor. In listing faunal distribution on a continental basis, data at the generic level would have been adequate, but to make use of the specific listings, distribution must be known in more detail; herein lies a major weakness of the book.

The size of the time steps used are important in any faunal analysis. Here, the authors have wisely chosen the "mammal age" as their basic unit. Although these are very uneven in time span (ranging from 0.5 to 9 million years), they are widely recognized and internationally agreed intervals. The text includes locality lists for each epoch, extensive bibliographies and, for most epochs, detailed justification for the choice of boundaries, even though the listings are based on the ages and not on the epochs.

In the final chapter the authors tabulate the distributional data for North American and European mammalian genera and families, providing graphs of the turnover rates and voicing (very softly) a few inter-

pretations and speculations.

Quite obviously this is a book to be referred to, not one to read. Nonetheless the text is quite understandable and in places even entertaining. There is a fondness for the second person (singular or plural?); *you* are treated to discussions on the biogeography, biogeography and palaeogeography of fossil mammals. The maps are not wholly successful, however — no scale, no latitude or longitude, often much wasted space — and very few of them are based on palaeogeographical maps. There is a scattering of fossil illustrations, inserted one feels at the publisher's request; nearly all are from publications of the University of California Press, which at least has the merit of maintaining consistently high standards. The number of misprints is smallish and these are mostly trivial and obvious.

There has been no work for the past quarter of a century which has attempted to record the mammalian palaeofaunas of the world, and none at any time which has done the job in such painstaking detail. On these grounds alone we must be grateful to the authors for providing us with a highly reliable work, an indispensable reference source and a monumental data base from which to test our theories, be they evolutionary, faunal or geographical. □

R.J.G. Savage is Professor of Vertebrate Palaeontology at the University of Bristol.

Back to the ocean

R.W. Girdler

The Ocean Basins and Margins: Vol. 6, The Indian Ocean.

Edited by Alan E.M. Nairn and Francis G. Stehli.

Plenum: 1982. Pp.776. \$85, £59.50.

NINETEEN eighty-three marks the fiftieth anniversary of the Anglo-Egyptian *Mabahiss-John Murray* expedition to the Indian Ocean, the event being celebrated in Alexandria this September. The sixth volume of this series on ocean basins and margins thus makes a timely appearance. There have of course been earlier and more recent expeditions to the region; the first was by *H.M.S. Challenger* in the 1870s, while in the 1960s there was the international Indian Ocean expedition in which some 40 ships from 13 nations took part.

The Indian Ocean best exemplifies continental drift. It was essentially formed by the break-up and drifting apart of Antarctica, Africa, India and Australasia. India drifted fastest, fairly rocketing north at nearly 10 cm yr⁻¹, crashing against Asia some 50 Myr ago and resulting in the evolution of the Himalayas. The Seychelles stand as sialic relics indicating that the continental break-up was not always clean and

making it easy to visualize continental drift, the islands being left behind as the major landmasses dispersed.

Somehow this weighty volume fails to capture the excitement of the exploration of the Indian Ocean and of the application of plate tectonics in the area. There are two notable exceptions — a comprehensive account of the aseismic ridges, spreading centres and basins by R. Schlich, and a fascinating review of the structure, tectonics and history of the north-eastern Indian Ocean by J.R. Curran and co-workers. The new charts and tectonic maps in these chapters are now very detailed, but it is remarkable how the 1965 physiographic diagram of the Indian Ocean by Heezen and Tharp stands the test of time. I happened to have been with Marie Tharp at the Lamont Geological Observatory the day she discovered the data were better fitted if all the fracture zones trended north-north-east instead of west-east as previously found in the Atlantic Ocean. This caused elation and dismay; elation at the discovery, dismay at the thought of having to re-do the whole chart! Now this all fits most beautifully into the plate tectonic picture.

The remaining chapters take us clockwise around the margins of the Indian Ocean. Starting with south-east Africa, the Somali Basin and east Africa, diversions into the Red Sea and Gulf of Aden, then the Persian Gulf, and follow on around the Indian continental margins, south-east Asia, western Australia, southern Australia and finally the Antarctic margin. The theme is broken by one chapter on oceanic islands and another on Madagascar. Apart from sialic fragments such as the Seychelles there are many volcanic islands in the Indian Ocean, and so I was hoping to discover from the contribution on oceanic islands how these have formed and how they are related to the sea-floor-spreading history, hot spots and so on. Alas, we find but an encyclopaedic account of the geology of the islands with few ideas of why they are there.

There are no introductory or concluding chapters and surprisingly, in view of the interests of the editors, no chapter on the palaeomagnetism of the continents surrounding the Indian Ocean to complement the accounts of the sea-floor-spreading history and plate tectonic setting. There are very comprehensive reference lists, but few citations post-1979.

Excitement may be missing, but the book certainly provides the reader with a feast of information. For those looking for such detail on the Indian Ocean and its margins this book will be of great value; for those wanting an overview of tectonic history and the evolutionary processes, the many papers which have appeared recently will be more useful. □

R.W. Girdler is Reader in Geophysics in the School of Physics, University of Newcastle upon Tyne.

**Mesozoic Mammals*. By J.A. Lillegraven, Z. Kielan-Jaworowska and W.A. Clemens. (University of California Press, 1979). *Pleistocene Mammals of Europe*. By B. Kurtén. (Weidenfeld & Nicholson, 1968). *Pleistocene Mammals of North America*. By B. Kurtén and E. Anderson. (Columbia University Press, 1980).

Not the end of the dinosaurs

Alan J. Charig

The Great Extinction: What Killed the Dinosaurs and Devastated the Earth?

By Michael Allaby and James Lovelock.
Doubleday/Secker and Warburg: 1983.
Pp.189. \$13.95, £10.95.

THE MAIN title of this semi-popular book is misleading, for the organic extinctions at the end of the Cretaceous are no more than a secondary theme. Its primary purpose, no doubt, is to demonstrate that approximately 65 million years ago the Earth was struck by a "planetesimal" with a diameter of 10 or 11 kilometres, a mass of 10^{11} or 10^{12} tons and a velocity of 20 kilometres per second; in other words, a meteorite "much larger and heavier than Mount Everest, made from solid rock and metal, approaching the Earth at about 20 times the speed of a high-velocity bullet from a modern army rifle". It came in at a steep angle, fell in the North Atlantic and disintegrated on impact, releasing as much energy as the detonation of 10^{14} tons of TNT and producing a huge crater. Its several effects were the causes of the end-of-Cretaceous extinctions.

Allaby and Lovelock's scenario differs from the well-known ideas of Alvarez *et al.* (*Science* 208, 1095; 1980) in three important particulars. First, it is more specific in postulating quite definitely that the planetesimal fell in the ocean rather than on land (Alvarez was indifferent on that point); conversely, it is less specific in suggesting that the green plants could have been killed not only by a temporary obliteration of sunlight but also by the complete masking of the leaves by dust. Thirdly, it proposes that a major cause — if not the major cause — of organic extinction was "air pollution on the heroic scale"; the causal agents produced by the planetesimal impact included "acid rain" (leading to an intolerably low pH of the soil water), the eutrophication of coastal waters and estuaries, and high concentrations of chlorine, of hydrogen cyanide and of compounds of certain toxic elements which are relatively abundant in extraterrestrial matter, such as osmium. In pursuing their thesis the authors touch upon a battery of scientific disciplines, both physical and biological, man's anthropomorphic attitude towards animals and his attitude towards future catastrophes.

Occasionally the style is lively and informative, couched at just the right level for a semi-popular book; sometimes, however, it is so dramatic as to verge on the sensational, and elsewhere the short staccato sentences are reminiscent of the front pages of the tabloid press. The chapter headings too have a journalistic

flavour, likewise the totally irrelevant and inconsequential paragraph concerning (*inter alia*) the scandalous relationships of Lord Byron. Altogether the result is much too diffuse for a book on science, even for a popular book. It is difficult to distinguish fact from speculation or important conclusions from the less important; indeed I am still uncertain as to what the authors regard as the principal immediate cause of the organic extinctions.

Several recent articles, alas, have expressed grave doubts on various aspects of the impact theories. Those aspects include the contemporaneity of the iridium anomalies at different places, their correlation with an extraterrestrial event, the suddenness and contemporaneity of the K-T extinctions, and likewise their correlation with a single catastrophe. Meanwhile my original interest in Allaby and Lovelock's book and my confidence in their reliability have been extinguished altogether by a careful reading of those parts which fall within my own field of specialization.

The authors' knowledge of dinosaurs and of fossil reptiles in general is pathetically inadequate, obviously derived from textbooks at least twenty years out of

date, and their arguments are often fallacious and based on misconceptions. Elementary howlers abound. For example, clay is *not* a "soil"; fossil reptiles *have* been found in Antarctica, from 1968 onwards; archosaurs were *not* primitively bipedal; ichthyosaurs, mosasaurs and pterosaurs are *not* dinosaurs; and leathery and green turtles are *not* confined to the Southern Hemisphere, being commoner in the Northern.

Their grasp of ecology too is sadly deficient, especially where it concerns the relationship between size, energy expenditure, reproduction and survival, and also the ability of the fauna to survive indefinitely without the input of photo-synthetic producers into the food web (apparently they fed on waste material and on each other!). Worst of all, betraying the authors' ignorance of the basics of systematics, was the absurd question (p.163): "Is not a homoiothermic, viviparous reptile, adapted to life in a seasonal climate, but a mammal by another name?" That, for me, was the last straw.

Alan J. Charig is Chief Curator of Fossil Amphibians, Reptiles and Birds at the British Museum (Natural History).

Odd conformation

Thomas E. Creighton

Protein Folding.

By Charis Ghéllis and Jeannine Yon.
Academic: 1982. Pp.562. \$74.50, £49.20.

THE complexity of the problem of how proteins acquire their folded conformations can be illustrated most graphically by calculating the average time required for a random polypeptide chain to sample all of its possible conformations. Even conservative estimates for relatively short polypeptide chains are many orders of magnitude longer than the age of the Universe. Yet proteins can fold within seconds or minutes, so the process must be non-random and directed in some way. It has been the subject of extensive theoretical and experimental studies, but there is still little consensus as to even its broad, general features.

A brave attempt to summarize the topic is made in this volume. A fresh, objective and critical approach might have been expected, since the authors are relatively new to the field (most of their own work appears to be unpublished), but the subject has proved too intractable for them. The treatment is quite comprehensive — often uncomfortably so for those in the field, for many of their past errors are repeated here, with remarkable naivety and with no qualification or correction. For example, the discussion of folding kinetics is introduced with a lengthy account (pp.329–350) of the purportedly classic theoretical analysis,

even though it was shown eight years ago to be inapplicable because it mistakenly assumed the unfolded state to be kinetically homogeneous. The disproved conclusions of that analysis (pp.362, 363, 366) and others (pp.350–355) are given with no qualification.

The material presented is often of dubious choice. The discussion of stabilizing interactions in proteins makes no mention of the classic experimental measurements on hydrogen bond stability by Klotz and co-workers, and those on hydrophobicity by Nozaki and Tanford are described in a mere seven lines. Instead, the discussion is based almost entirely on some questionable theoretical calculations. Much of the experimental data presented on protein structure and stability is twenty years old and has been superseded; in particular, the landmark calorimetric measurements by Privalov are virtually ignored. Finally, the number of grammatical and typographical errors is inexcusable in a book selling at this price.

This volume may be a useful compendium of references but most potential readers would be better served by the eleven extensive reviews and one symposium volume on the subject that have appeared in the past eight years. Collectively they do not give a coherent picture of protein folding; but at least they are internally consistent, perpetuate fewer errors and present different points of view. □

Thomas E. Creighton is a member of the scientific staff at the MRC Laboratory of Molecular Biology, Cambridge.